The Student's Physics

MODERN PHYSICS

BLACKIE & SON LIMITED
66 Chandos Place, LONDON
17 Stanhope Street, GLASGOW

BLACKIE & SON (INDIA) LIMITED
103/5 Fort Street BOMBAY

BLACKIE & SON (CANADA) LIMITED
TORONTO

MODERN PHYSICS

BY

H. A. WILSON

M.Sc (Leeds), M.A (Camb), D Sc (Lond), F R S.

Professor of Physics, Rice Institute, Houston, Texas, U S A.

Formerly Fellow of Trinity College, Cambridge

THIRD EDITION

BLACKIE & SON LIMITED

LONDON AND GLASGOW

First issued 1928
Reprinted with appendix 1930
Second edition 1937
Reprinted, with corrections, 1944
Reprinted 1946
Third edition 1948

Printed in Great Britain by Blackie & Son, Ltd., Glasgow

PREFACE TO FIRST EDITION

This book is based largely on lectures which I have given to students taking an Honours Course in Physics during recent years. It is intended to serve as an introduction to the treatises and papers on particular branches of the subject to several of which reference is made at the end of each chapter. It is impossible for any student to acquire a complete knowledge of all branches of physics, and in this book I have tried only to give a concise but intelligible account of no more than a serious student of physics ought to be familiar with when he begins to specialize on some particular branch.

The rapid changes which are now taking place in several of the subjects discussed make it difficult to write a consistent and up-to-date account of them.

Very few references to original papers have been given. These will be found in the treatises on special branches mentioned at the end of each chapter.

No attempt has been made to assign all the facts and theories discussed to their original authors. This has been done only in a few cases. The particular facts and theories selected for discussion in such a book depend largely on the knowledge and experience of the writer, and it is impossible to avoid many omissions of important results some of which another author would very likely have considered it essential to include.

My best thanks are due to Dr. John Dougall for his careful reading of the proofs and for many valuable suggestions. Also to Sir Ernest Rutherford and Mr. Blackett of the Cavendish

Laboratory for permission to reproduce the beautiful photographs of α-ray tracks.

H. A. WILSON.

RICE INSTITUTE, HOUSTON, TEXAS,
May, 1928.

PREFACE TO SECOND EDITION

For the new edition, the work has been subjected to a thorough revision, and much new matter has been added. In particular, the chapter on the Quantum Theory has been almost entirely rewritten, and a new chapter has been added on Atomic Nuclei.

H. A. WILSON.

RICE INSTITUTE, HOUSTON, TEXAS,
August, 1937.

PREFACE TO THIRD EDITION

For the new edition the chapters on Magnetism, Quantum Theory and Atomic Nuclei have been partly rewritten with additions necessary to bring them up to date. An account of superconductivity has been added to the chapter on Electron Theory. The new matter includes an account of nuclear fission, the atomic bomb, and the meson theory of nuclear forces.

H. A. WILSON.

RICE INSTITUTE, HOUSTON, TEXAS,
November, 1947.

CONTENTS

CHAPTER I

THE ELECTRON THEORY

CHAPTER II

THEORIES OF MAGNETISM

CHAPTER III

THERMIONICS

CHAPTER IV

PHOTO-ELECTRICITY

CHAPTER V

THE QUANTUM THEORY

CHAPTER VI

THE CRITICAL POTENTIALS OF ATOMS

CHAPTER VII

X-RAYS AND γ-RAYS

CHAPTER VIII

OPTICAL SPECTRA

CHAPTER IX

CATHODE RAYS, β-RAYS, AND α-RAYS

CHAPTER X

POSITIVE RAYS

CHAPTER XI

RADIOACTIVE TRANSFORMATIONS

CHAPTER XII

ATOMIC NUCLEI

CHAPTER XIII

GASEOUS IONS

CHAPTER XIV

THE MOTION OF ELECTRONS IN GASES

CHAPTER XV

THE ELECTRICAL CONDUCTIVITY OF FLAMES

CHAPTER XVI

THE POSITIVE COLUMN AND NEGATIVE GLOW

CHAPTER XVII

ATMOSPHERIC ELECTRICITY

CHAPTER XVIII

SPECIAL RELATIVITY

CHAPTER XIX

GENERAL RELATIVITY AND GRAVITATION

CHAPTER XX

MATHEMATICAL NOTES

MODERN PHYSICS

CHAPTER I

The Electron Theory

1. Electrons and Positive Nuclei.

The principal electrical properties of matter are specific inductive capacity, conductivity, and magnetic permeability. When the numerical values of these quantities for the bodies present in a system are known, then the electrical phenomena to be expected in the system may be worked out. For most practical purposes it is sufficient to know that different forms of matter have these properties in greater or less degree, and it is not necessary to attempt to explain why this is so.

For scientific purposes, however, it is desired to explain the properties of matter, and to do this it is necessary to consider the nature of the ultimate particles composing material bodies. The properties, such as those just mentioned, which can be measured, are properties of matter in bulk or of bodies containing enormous numbers of atoms. They represent average values taken over volumes very large compared with the volume of one atom.

According to the electron theory material bodies contain enormous numbers of minute particles of negative electricity which are all equal. These atoms of negative electricity are called electrons. Electrically neutral bodies contain equal amounts of positive and negative electricity, and the positive electricity is also supposed to consist of minute particles which are the nuclei of the atoms.

The electrons and positive nuclei are supposed to be so small that even in the densest solids such as gold they only occupy a very minute fraction of the space. Material bodies therefore are merely space with particles of electricity here and there separated by distances very large compared with the dimensions of the particles. The space between the particles is not empty, for it is filled with the electrical and magnetic fields excited by the particles. Thus if we could select a point at random

inside a material body and examine the state of the space at that point we should almost always find nothing except an electric and a magnetic field.

The point would be in a vacuum and the specific inductive capacity, conductivity, and permeability at the point would have the usual values for a perfect vacuum. If the point happened to be inside an electron or a positive nucleus, then we may suppose that there would be a certain density of electric charge at the point in addition to the electric and magnetic fields.

2. Fundamental Electromagnetic Equations of the Electron Theory.

The electromagnetic equations for the interior of matter according to the electron theory are the equations for space containing nothing but electricity. The only quantities which are required in these equations are density of electricity or charge per unit volume, electric field strength, and magnetic field strength. There is no magnetism on this theory, and magnetic fields are supposed to be produced only by the motion of electricity. Following H. A. Lorentz, to whom the development of the electron theory is largely due, we shall use Heaviside's rational units and suppose that the force between two charges e_1 and e_2, at rest at a distance r apart, is equal to $e_1e_2/4\pi r^2$. This amounts to taking for the unit charge a quantity of electricity equal to the usual electrostatic unit divided by $\sqrt{4\pi}$. The field strength will as usual be the force on a unit charge, so that the field strength due to a charge e at rest is $e/4\pi r^2$. The number of lines of force crossing unit area, drawn perpendicular to the field, will, also as usual, be taken equal to the strength of the field. The number of lines of force ending on a negative charge e or starting from a positive charge e is therefore $4\pi r^2 \times \frac{e}{4\pi r^2}$, which is equal to e. The number of lines of force starting in unit volume is therefore equal to the density of charge, which is expressed by the equation

$$\operatorname{div}^* \mathbf{F} = \rho,$$

where $\mathbf{F}$ denotes the field strength and ρ the density of charge. The force between two magnetic poles will be taken to be $m_1m_2/4\pi r^2$, so that the field strength due to a pole m is $m/4\pi r^2$. Since we suppose that there is no magnetism all lines of magnetic force must be closed curves, so that the number of lines which end in unit volume is zero, and $\operatorname{div} \mathbf{H} = 0$, where $\mathbf{H}$ denotes the magnetic field strength.

The current density, that is the current per unit area drawn perpendicular to the direction of the current, at a point between the electrons and positive nuclei will be simply the displacement current

* The vector notation and nomenclature used here is explained at p. 398

and so will be equal to $\partial\mathbf{F}/\partial t$. At a point inside an electron or a nucleus where the density of charge is ρ there will be also, if the charge is moving, a convection current density equal to $\rho\mathbf{V}$, where $\mathbf{V}$ is the velocity of the electricity. The total current density at any point is therefore the resultant or vector sum of $\partial\mathbf{F}/\partial t$ and $\rho\mathbf{V}$, or $\frac{\partial\mathbf{F}}{\partial t}+\rho\mathbf{V}$. $\mathbf{F}$ and ρ being expressed in the Heaviside electrostatic units, the unit of current density will of course be one Heaviside electrostatic unit of charge per square centimetre per second. The magnetic field $\mathbf{H}$ is excited by the current according to the well-known relation that the work required to take a unit pole once round a current is equal to 4π times the current, the field and current being expressed in ordinary electromagnetic units. In vector notation, curl $\mathbf{H}=4\pi\mathbf{i}$, where $\mathbf{i}$ is the current density. If $\mathbf{i}$ is expressed in Heaviside electrostatic units and $\mathbf{H}$ in Heaviside electromagnetic units this becomes curl $\mathbf{H}=\mathbf{i}/c$, where c is the velocity of light, so that we have

$$\text{curl } \mathbf{H}=\frac{1}{c}\left(\frac{\partial\mathbf{F}}{\partial t}+\rho\mathbf{V}\right).$$

In the same way the well-known relation between the induced electromotive force round a closed curve and the rate of variation of the magnetic field through the curve may be written

$$\text{curl } \mathbf{F}=-\frac{1}{c}\frac{\partial\mathbf{H}}{\partial t}.$$

The four equations

$$\text{div } \mathbf{H}=0,\quad \text{div } \mathbf{F}=\rho,$$
$$\text{curl } \mathbf{H}=\frac{1}{c}\left(\frac{\partial\mathbf{F}}{\partial t}+\rho\mathbf{V}\right),\quad \text{curl } \mathbf{F}=-\frac{1}{c}\frac{\partial\mathbf{H}}{\partial t},$$

are the fundamental electromagnetic equations of the electron theory. If $\rho=0$, they reduce to the electromagnetic equations for space devoid of ponderable matter, or ether, as it is sometimes called.

These equations theoretically enable $\mathbf{H}$ and $\mathbf{F}$ to be calculated when ρ and $\mathbf{V}$ are given throughout space as functions of the time t.

The velocity $\mathbf{V}$ of the electricity is supposed to be measured relatively to the material system on which the observer is. In all ordinary cases $\mathbf{V}$ will be the velocity relative to the earth or to the laboratory in which the observer is working. The field strengths $\mathbf{F}$ and $\mathbf{H}$ also are the fields as measured by an observer relative to the material system on which he is working, which is usually the earth. It is found that the motion of the earth has no observable influence on electromagnetic phenomena so that it is customary in electrical experiments to regard the laboratory as at rest.

3. Remarks on the Fundamental Equations and on the Fields.

The fundamental electromagnetic equations indicate that each charge produces a field as though the other charges were not present Thus if all the charges are at rest and there is no magnetic field anywhere, then the electric field is given by div $\mathbf{F} = \rho$. This makes the field due to a charge e equal to $e/4\pi r^2$, and the potential $e/4\pi r$. The resultant field at any point is then the resultant of the fields due to every charge, each charge giving a field $e/4\pi r^2$. The total potential at any point is simply the sum of the potentials due to all the charges present. This means that the presence of material bodies in no way modifies the fields due to charges Any charge excites the same field inside a material body as in empty space. This is not surprising when we remember that material bodies are almost entirely empty space, since the electrons and nuclei occupy only a minute fraction of the space even in the densest substances It is possible that if a body could be obtained with such a high density that the electrons and nuclei in it filled an appreciable fraction of its volume, the fields produced in it would differ appreciably from those produced by the same charges in empty space.

The fundamental equations of the electron theory lead to a fairly satisfactory explanation of the electrical properties of material bodies containing very large numbers of electrons and nuclei They fail, however, to explain the properties of single atoms It is possible that they represent results which are true on the average over large numbers of electrons and time intervals not too short, but which are not true for very small numbers of electrons and very short time intervals. For example, the equation div $\mathbf{F} = \rho$ may really only be true when ρ is the average density of charge over a volume containing a very large number of electrons and nuclei. However, in the electron theory we assume that this equation is true when ρ is the charge in an indefinitely small volume divided by that volume The volume considered may be small compared with the volume of a nucleus even. We assume that the results of large-scale experiments on bodies containing enormous numbers of electrons and nuclei are true for microscopic phenomena even inside electrons.

The fundamental equations give the electric and magnetic fields excited in the surrounding space by electric charges. It is important to remember that electric and magnetic fields are not directly observed in any experiments Only phenomena in material bodies are observed. The existence of fields in the space surrounding charges is assumed because phenomena can be conveniently described or explained by means of this assumption. In the electron theory we assume the existence of these fields but we do not attempt to explain how they are produced or what they consist of. We may if we like regard them as merely auxiliary mathematical quantities introduced into the theory for convenience in attempting to describe phenomena. Most physicists, however, believe that these fields really exist. It has been suggested that electric and magnetic fields are modifications of the ether, a medium filling all space. This hypothesis is not of much use; it is sufficient to suppose that the charges excite the fields in the surrounding space. If all space is filled with ether, then ether and space are the same thing, and we may as well regard electric and magnetic fields as modifications of space or as merely existing in space. The distinction between ether and space will never amount to anything until some method of removing the ether from a portion of space is discovered, and we have no reason to hope that anything of the sort will ever be possible.

The electric and magnetic field strengths $\mathbf{F}$ and $\mathbf{H}$ are vectors, and therefore the electromagnetic equations necessarily represent relations which are geometrically possible in vector fields. Thus if $\mathbf{A}$ denotes a vector the components of which are continuous throughout a vector field, and div $\mathbf{A} = 0$ everywhere, then we can always find a vector $\mathbf{B}$ such that curl $\mathbf{B} = \mathbf{A}$. Hence, for example,

it is always possible to find an electric field $\mathbf{F}$ such that curl $\mathbf{F} = -\frac{1}{c}\dot{\mathbf{H}}$, since div $\mathbf{H} = 0$, and therefore div $\dot{\mathbf{H}} = 0$ The fact that a varying magnetic field excites an electric field is nevertheless an experimental result which is not even suggested by the geometrical properties of vectors.

4 Calculation of the Magnetic Field Strength. The Vector Potential.

In order to see how to calculate the field strengths $\mathbf{F}$ and $\mathbf{H}$ at any point, when ρ and $\mathbf{V}$ are supposed given as functions of the time t throughout the surrounding space, we first eliminate $\mathbf{F}$ from the fundamental equations. To do this we use the well-known vector equation *

$$\text{curl curl } \mathbf{A} = \text{grad div } \mathbf{A} - \Delta\mathbf{A},$$

which is true for any vector $\mathbf{A}$. Hence, since div $\mathbf{H} = 0$, we have

$$\text{curl curl } \mathbf{H} = -\Delta\mathbf{H} = \frac{1}{c}\text{ curl } (\dot{\mathbf{F}} + \rho\mathbf{V}).$$

But curl $\mathbf{F} = -\frac{1}{c}\dot{\mathbf{H}}$, so that curl $\dot{\mathbf{F}} = -\frac{1}{c}\ddot{\mathbf{H}}$, and therefore

$$\Delta\mathbf{H} - \frac{1}{c^2}\ddot{\mathbf{H}} = -\frac{1}{c}\text{ curl } \rho\mathbf{V}.$$

If V is any quantity and $\Delta V = \omega$, then we have as the solution of this differential equation

$$V_P = -\frac{1}{4\pi}\int\frac{\omega\, dS}{r},$$

where V_P is the value of V at a point P, r is the distance from an element of volume dS to P, and ω the value of ω in the element dS. The integral is supposed extended over all the space around P in which ω differs from zero. If V and ω are vectors then this equation is equivalent to three equations, one for each component, e.g.

$$V_x = -\frac{1}{4\pi}\int\frac{\omega_x\, dS}{r},$$

where V_x and ω_x denote the x components of $\mathbf{V}$ and $\boldsymbol{\omega}$.

If instead of $\Delta V = \omega$, we have

$$\Delta V - \frac{1}{c^2}\ddot{V} = \omega,$$

then the solution is

$$V_P = -\frac{1}{4\pi}\int\frac{[\omega]\, dS}{r},$$

where $[\omega]$ stands for the value of ω, not at the time t at which the value

* See p. 398

of V at the point P is V_P, but at the earlier time $t - r/c$. The effect of the term $-\dot{V}/c^2$ is to make the field excited by ω travel out from ω with the velocity c.

The solution of the equation obtained above for $\mathbf{H}$ is therefore

$$\mathbf{H} = \frac{1}{4\pi c}\int\frac{[\text{curl } \rho\mathbf{V}]}{r}dS.$$

Since the calculation of $\mathbf{H}$ by means of this equation is rather complicated it is usual to introduce an auxiliary mathematical quantity $\mathbf{a}$, called the vector potential, which is given by the equation

$$\mathbf{a} = \frac{1}{4\pi c}\int\frac{[\rho\mathbf{V}]dS}{r},$$

so that $\mathbf{H} = \text{curl } \mathbf{a}$. Here again the square brackets mean that the value of $\rho\mathbf{V}$ in the element of volume dS is to be taken at the time $t - r/c$ in order to get the value of $\mathbf{a}$ at the time t.

5. Calculation of the Electric Field Strength. The Scalar Potential.

If we eliminate $\mathbf{H}$ from the electromagnetic field equations in the same way as $\mathbf{F}$ was eliminated we get

$$\Delta\mathbf{F} - \frac{1}{c^2}\ddot{\mathbf{F}} = \text{grad } \rho + \frac{1}{c^2}\frac{\partial}{\partial t}(\rho\mathbf{V}).$$

The solution of this is

$$\mathbf{F} = -\frac{1}{4\pi}\int\frac{\left[\text{grad } \rho + \frac{1}{c^2}\frac{\partial}{\partial t}(\rho\mathbf{V})\right]}{r}dS.$$

Instead of calculating $\mathbf{F}$ by means of this vector equation we may introduce an auxiliary mathematical quantity ϕ, called the scalar potential, which is given by

$$\phi = \frac{1}{4\pi}\int\frac{[\rho]}{r}dS,$$

and $\mathbf{F}$ is then given by the equation

$$\mathbf{F} = -\frac{1}{c}\dot{\mathbf{a}} - \text{grad } \phi,$$

since
$$\frac{1}{4\pi}\int\frac{\left[\frac{1}{c^2}\frac{\partial}{\partial t}(\rho\mathbf{V})\right]}{r}dS = \frac{1}{c}\dot{\mathbf{a}}.$$

When ρ and $\mathbf{V}$ are known as functions of the time at all points in the

space around a point P, ϕ and $\mathbf{a}$ can be calculated at any point near P, and then $\mathbf{H}$ and $\mathbf{F}$ are given by

$$\mathbf{H} = \text{curl } \mathbf{a}, \quad \mathbf{F} = -\frac{1}{c}\dot{\mathbf{a}} - \text{grad } \phi.$$

6. Examples of the Use of the Equations.

As an example of the use of these equations we will work out the fields due to a single element of volume in which ρ and $\mathbf{V}$ are supposed constant For this element of volume let $\rho dS = e$, so that $\varphi = e/4\pi r$ and $\mathbf{a} = e\mathbf{V}/4\pi cr$.

Let the element of volume considered be at the origin of co-ordinates so that $r^2 = x^2 + y^2 + z^2$, and let the velocity $\mathbf{V}$ be along the x axis, so that $V_x = V$, $V_y = 0$, $V_z = 0$. Then we have

$$a_x = \frac{eV}{4\pi cr}, \quad a_y = 0, \quad a_z = 0, \quad \text{and } \dot{\mathbf{a}} = 0.$$

Hence

$$H_x = (\text{curl } a)_x = 0,$$

$$H_y = (\text{curl } a)_y = \frac{\partial}{\partial z}\left(\frac{eV}{4\pi cr}\right) = -\frac{eVz}{4\pi cr^3},$$

$$H_z = (\text{curl } a)_z = -\frac{\partial}{\partial y}\left(\frac{eV}{4\pi cr}\right) = \left(\frac{eVy}{4\pi cr^3}\right).$$

The resultant magnetic field H is therefore equal to $\frac{eV}{4\pi cr^3}\sqrt{z^2 + y^2}$, or, if we put $\sqrt{z^2 + y^2} = r \sin\theta$, it is equal to $\frac{eV}{4\pi cr^2} \sin\theta$. Its direction is perpendicular to the plane containing r and $\mathbf{V}$, so that the lines of magnetic force are circles with their centres on the x axis and their planes perpendicular to the x axis. The electric field $\mathbf{F} = -$ grad φ, since $\dot{\mathbf{a}} = 0$, which gives $\mathrm{F} = \frac{e}{4\pi r^2}$.

If we suppose that the element of volume considered contains equal amounts of positive and negative electricity so that $\rho = 0$ in it, but that the positive electricity only has the velocity V, then we get

$$\varphi = 0, \quad \text{and } a_x = \frac{eV}{4\pi cr},$$

so that the element produces the resultant magnetic field $H = \frac{eV \sin\theta}{4\pi c r^2}$, but no electric field. The element of volume is then equivalent to an element of a conductor carrying a current, and the formula we have obtained for H gives Ampere's formula for the magnetic field due to a current element, since eV is equivalent to ids, where i is the current and ds the length of the element.

As another example, suppose that an electron with charge e has been at rest at the origin from $t = -\infty$ to a time t_1, and that between t_1 and t_2 it moves a small distance away from the origin and then back again and then remains at rest. Up to t_1 the field will be $e/4\pi r^2$ along r, but when the electron begins to move the field will change and the change will move out from the charge with velocity c. If we describe *two* spheres with centres at the origin and radii $c(t - t_1)$ and $c(t - t_2)$, then outside the larger sphere the disturbance due to the motion will not have arrived, and inside the smaller sphere the disturbance will have passed by, so that the field will be $\frac{e}{4\pi r^2}$ along r except in the space between the two spheres. Inside

this space there will be a magnetic field due to the motion as well as an electric field Thus the motion of the electron produces a wave in the field which moves out from it with the velocity c.

7 Poynting's Theorem.

We shall as usual suppose that the energy density in the electromagnetic field is $\frac{1}{2}(\mathbf{F}^2 + \mathbf{H}^2)$, so that if $\mathbf{P}$ denotes the flow of energy through unit area in unit time, then

$$-\operatorname{div} \mathbf{P} = \frac{1}{2} \frac{d}{dt}(\mathbf{F}^2 + \mathbf{H}^2) = (\mathbf{F} \cdot \dot{\mathbf{F}}) + (\mathbf{H} \cdot \dot{\mathbf{H}}).$$

In a space where $\rho = 0$ we have

$$\dot{\mathbf{F}} = c \operatorname{curl} \mathbf{H}, \quad \text{and } \dot{\mathbf{H}} = -c \operatorname{curl} \mathbf{F},$$

so that we get

$$\begin{aligned} -\operatorname{div} \mathbf{P} &= c(\mathbf{F} \cdot \operatorname{curl} \mathbf{H}) - c(\mathbf{H} \cdot \operatorname{curl} \mathbf{F}) \\ &= -c \operatorname{div} [\mathbf{F} \cdot \mathbf{H}]. \end{aligned}$$

Hence we conclude that $\mathbf{P} = c[\mathbf{F} \cdot \mathbf{H}]$, a result first obtained by Poynting and known as Poynting's Theorem According to this the flow of energy is perpendicular to both the electric and magnetic fields and equal per unit area per unit time to $c\,\mathrm{FH} \sin\theta$, where θ is the angle between $\mathbf{F}$ and $\mathbf{H}$.

8. Electromagnetic Momentum.

We know that there is a force on a current in a magnetic field due to the interaction of the magnetic field of the current and the external field. In space where $\rho = 0$ the current density is $\dot{\mathbf{F}}$, so that there must be a force on the electromagnetic field equal to $\frac{1}{c}[\dot{\mathbf{F}} \cdot \mathbf{H}]$ per unit volume. Also there must be a similar force on a varying magnetic field in an electric field, for a varying magnetic field produces an electric field in exactly the same way that a current produces a magnetic field. This force will be $\frac{1}{c}[\mathbf{F} \cdot \dot{\mathbf{H}}]$ per unit volume.

The total force on the field is therefore $\frac{1}{c}[\dot{\mathbf{F}} \cdot \mathbf{H}] + \frac{1}{c}[\mathbf{F} \cdot \dot{\mathbf{H}}]$, which is equal to $\frac{1}{c}\frac{d}{dt}[\mathbf{F} \cdot \mathbf{H}]$ or $\dot{\mathbf{P}}/c^2$ per unit volume. Since force is equal to rate of change of momentum we conclude that the field has momentum equal to $\mathbf{P}/c^2$ per unit volume. This momentum is called the electromagnetic momentum of the field. Its existence was first pointed out by J. J. Thomson. We do not regard this momentum as momentum of the ether but as momentum of the electromagnetic field. Thus the field has energy and momentum and can move through space so that

it has the most essential properties of matter. We regard the field as excited in space by the charges and it is not necessary to introduce the idea of an ether into the discussion.

Since it appears that the momentum is equal to $\mathbf{P}/c^2$, we conclude that the momentum is due to the flux of energy $\mathbf{P}$, so that we may say that electromagnetic energy when moving has momentum. The momentum of matter may therefore be supposed to be simply the momentum of the energy it contains. For energy can be converted from one kind into another, so that if electromagnetic energy has momentum it is difficult to see how other kinds of energy could not have the same amount of momentum.

9. Momentum and Energy of Matter.

Matter perhaps contains other kinds of energy besides electromagnetic energy, but we shall suppose that all the energy in matter has the same momentum as electromagnetic energy. If we make this assumption we can easily find the momentum of any material system in terms of its energy.

Let $\bar{x}$ denote the x co-ordinate of the centroid of the energy in any system, so that

$$\bar{x} = \frac{\int Ex\,dS}{\int E\,dS},$$

where E is the energy density in the element of volume dS, and the integrals are supposed taken over the whole of the system. Putting $\mathcal{E} = \int E\,dS$ for the total energy we get

$$\mathcal{E}\bar{x} = \int Ex\,dS.$$

Differentiating this with respect to the time t we get

$$\mathcal{E}\dot{\bar{x}} = \int \dot{E}x\,dS;$$

but $\dot{E} = -\operatorname{div}\mathbf{P}$, so that

$$\mathcal{E}\dot{\bar{x}} = -\int x(\operatorname{div}\mathbf{P})\,dS.$$

Now $$x\operatorname{div}\mathbf{P} = \frac{\partial}{\partial x}(xP_x) + \frac{\partial}{\partial y}(xP_y) + \frac{\partial}{\partial z}(xP_z) - P_x,$$

and $$\int\left\{\frac{\partial}{\partial x}(xP_x) + \frac{\partial}{\partial y}(xP_y) + \frac{\partial}{\partial z}(xP_z)\right\}dS$$

is, by Green's Theorem, equal to a surface integral over a surface enclosing the volume S. If all the energy is inside this closed surface then

P is zero all over it, so that the integral must be equal to zero. Hence

$$\mathcal{E}\dot{\bar{x}} = \int P_x dS.$$

But $P_x = M_x c^2$, where M_x is the x momentum density, so that if $\mathcal{M}_x = \int M_x dS$ we get

$$\mathcal{M}_x = \frac{\mathcal{E}\dot{\bar{x}}}{c^2}.$$

Thus the momentum of any system is equal to its energy multiplied by the velocity of the centroid of its energy and divided by the square of the velocity of light.

10. Relations between the Energy, Momentum, Velocity, and Mass of a Particle.

For a small particle of any kind the velocity v of the particle may be put equal to the velocity of the centroid of its energy, so that for any particle

$$\mathcal{M} = \frac{\mathcal{E}v}{c^2}.$$

The mass m of the particle may be defined as the momentum $\mathcal{M}$ divided by the velocity v so that $m = \mathcal{E}/c^2$.

If the energy $\mathcal{E}$ and mass m when the particle is at rest or when $v = 0$ are denoted by $\mathcal{E}_0$ and m_0, then $m_0 = \mathcal{E}_0/c^2$. The kinetic energy T of the particle is then given by $T = \mathcal{E} - \mathcal{E}_0$ so that $T = c^2(m - m_0)$.

If a force f acts on such a particle along the direction in which it is moving, and we suppose that there is no loss of energy by radiation or otherwise, then we have

$$f\delta t = \delta\mathcal{M}$$

and

$$fv\,\delta t = \delta\mathcal{E} = c^2\,\delta m,$$

so that $v\,\delta\mathcal{M} = c^2\,\delta m$, or since $\mathcal{M} = mv$, $\mathcal{M}\,\delta\mathcal{M} = c^2 m\,\delta m$.

Integrating this gives $\mathcal{M}^2 = c^2m^2 + \text{const.}$, so that, since $m = m_0$ when $\mathcal{M} = 0$,

$$\mathcal{M}^2 = c^2(m^2 - m_0{}^2) = m^2v^2.$$

This gives $m = m_0/\sqrt{1 - v^2/c^2}$, so that since $\mathcal{M} = mv$ and $m = \mathcal{E}/c^2$, we get

$$\mathcal{E} = \frac{\mathcal{E}_0}{\sqrt{1 - v^2/c^2}}$$

and
$$\mathcal{M} = \frac{\mathcal{E}_0 v}{c^2\sqrt{1 - v^2/c^2}}.$$

The mass m is sometimes called the apparent mass of the particle. It appears that m increases with v and becomes infinite when $v = c$. It is therefore impossible for the velocity of a material system to be as great as the velocity of light c. The kinetic energy of the particle is

$$\mathcal{E} - \mathcal{E}_0 = \mathcal{E}_0\left\{\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right\}.$$

When v/c is very small this gives

$$\mathcal{E} - \mathcal{E}_0 = \mathcal{E}_0\left\{1 + \frac{v^2}{2c^2} - 1\right\} = \frac{\mathcal{E}_0 v^2}{2c^2},$$

or, since $m_0 = \mathcal{E}_0/c^2$, we have $\mathcal{E} - \mathcal{E}_0 = \frac{1}{2}m_0v^2$.

11. Relation of the Mass of an Electron to its Charge.

If the particle considered is an electron, m will be the mass of the electromagnetic field which it excites and which moves along with it, together with any additional mass which it may have. If the electron is merely an electric charge it may have no additional mass, but if it has some internal energy besides its electrical energy it will have some additional mass corresponding to this additional energy. In any case its mass should vary with its velocity in accordance with the expression found above for m, since this should hold for a particle of any kind. The experiments of Kaufmann, Bucherer, and others on the variation of the mass of electrons with their velocity have shown that the mass does vary approximately in accordance with the above formula. These experiments confirm the idea that momentum is due to flux of energy, but they give no information as to the constitution of electrons.

If we assume that an electron consists merely of a charge e uniformly distributed over the surface of a sphere of radius a, then its electrical energy when at rest is $\frac{e^2}{8\pi a}$, so that its mass m_0 should be $\frac{e^2}{8\pi a c^2}$. However, such a sphere of electricity would tend to fly apart owing to the repulsion

between the different parts of its charge To counteract this repulsion we may suppose that there is a tension inside it equal to the repulsion or to $\frac{1}{2}\left(\frac{e}{4\pi a^2}\right)^2$. This tension may indicate the presence of an amount of energy equal to the volume of the sphere multiplied by the tension or to $\frac{1}{2}\left(\frac{e}{4\pi a^2}\right)^2 \frac{4}{3}\pi a^3 = \frac{e^2}{24\pi a}$.

According to this the total energy will be

$$\frac{e^2}{24\pi a} + \frac{e^2}{8\pi a} = \frac{e^2}{6\pi a},$$

so that the mass

$$m_0 = \frac{e^2}{6\pi a c^2}.$$

From the known values of e/m_0 and e we can calculate a, which comes out about 10^{-13} cm. We have reasons for believing that electrons are very small, but there is of course no justification for the idea that an electron is a sphere of electricity. Electricity as we know it consists of electrons, and a part of an electron if it could be examined would very likely be found to have properties quite different from those of electricity. We might as well suppose that a part of an atom of, say, helium would have the properties of a helium atom as that part of an electron or atom of electricity would have the properties of an electron. The density of electricity ρ which appears in the electromagnetic equations is supposed in the electron theory to be the charge in an element of volume divided by the volume of the element, and it is assumed that the element of volume considered may be inside an electron and small compared with the electron. Such assumptions are not really justified; all experiments relate to volumes enormously large compared with electrons, and we are really only justified in supposing that ρ in the electromagnetic equations stands for such a quantity as $\frac{ne}{V}$, where n is the number of electrons, each having a charge e, contained in a small volume V.

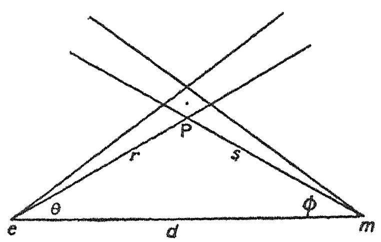


FIG. 1

12. Force on a Charge moving in a Magnetic Field.

The force on a charge when moving in a magnetic field can be obtained by considering the electromagnetic momentum due to a charge and a magnetic pole. Consider a charge e at a distance d from a pole of strength m (fig. 1).

Consider a point P at a distance r from e and s from m. The electric field at P due to e is $\frac{e}{4\pi r^2}$, and the magnetic field is $\frac{m}{4\pi s^2}$. The electromagnetic momentum at P per unit volume is therefore $\frac{em\sin(\theta+\phi)}{16\pi^2 r^2 s^2 c}$, where θ is the angle Pem and ϕ the angle Pme. As element of volume let us take a ring element described by the element of area bounded by the lines θ, $\theta+d\theta$, ϕ, $\phi+d\phi$, when the plane containing θ and ϕ rotates once round the line em as axis. The element of area is equal to $\frac{rs\,d\theta\,d\phi}{\sin(\theta+\phi)}$, so that the ring element of volume is equal to $2\pi r\sin\theta\frac{rs\,d\theta\,d\phi}{\sin(\theta+\phi)}$. The angular momentum in the ring element about the line em as axis is therefore

$$\frac{s\sin\phi\,em\sin(\theta+\phi)\,2\pi r\sin\theta\,rs\,d\theta\,d\phi}{16\pi^2 r^2 s^2 c\sin(\theta+\phi)}=\frac{em\sin\theta\sin\phi\,d\theta\,d\phi}{8\pi c}.$$

The total angular momentum will be got by integrating with respect to θ from $\theta=0$ to $\theta=\pi-\phi$, and then with respect to ϕ from $\phi=0$ to $\phi=\pi$. This gives for the total angular momentum $\frac{em}{4\pi c}$, a result first obtained by J. J. Thomson, to whom this interesting method of calculating the force on a charge moving in a magnetic field is due. Now suppose the charge is moving with velocity v in a direction making an angle ψ with the magnetic field. In a time dt the line joining e and m will turn through an angle $\frac{v\sin\psi\,dt}{d}$, so that the angular momentum is changed in direction, and this requires the addition of angular momentum about an axis in the plane containing e, m, and v, equal to

$$\frac{emv\sin\psi\,dt}{4\pi cd}.$$

If a force f acts on the charge in a direction perpendicular to the plane containing e, m, and v, and an equal and opposite force on the pole, then we may equate the couple to the rate of change of the angular momentum, so getting

$$fd=\frac{emv\sin\psi}{4\pi cd}.$$

Hence

$$f=\frac{emv\sin\psi}{4\pi cd^2}.$$

If we put $H=\frac{m}{4\pi d^2}$, this becomes

$$f=\frac{Hev\sin\psi}{c},$$

so that for the force on a unit charge moving with velocity $\mathbf{V}$ we get $\frac{1}{c}[\mathbf{V}\cdot\mathbf{H}]$. The total force on unit charge in the electromagnetic field is therefore the resultant or vector sum of the force due to the electric field and that due to the magnetic field, or

$$\mathbf{F}+\frac{1}{c}[\mathbf{V}\cdot\mathbf{H}].$$

This equation is of equal importance with the other fundamental electromagnetic equations.

13. **Force on a Moving Electron.**

When an electron having the charge e is moving with velocity $\mathbf{V}$ in an electromagnetic field, the force on it due to the external fields $\mathbf{F}$ and $\mathbf{H}$ will be

$$e\left(\mathbf{F}+\frac{1}{c}[\mathbf{V}\cdot\mathbf{H}]\right),$$

provided we suppose that $\mathbf{F}$ and $\mathbf{H}$ do not vary appreciably in the small space occupied by the electron. In addition to this force on the electron there will be the force on it due to its own field. If the mass of the electron is zero then the total force on it must be zero, but if it contains some internal energy and so has some mass, then the total force will be equal to the rate of change of its momentum. We do not know how much internal energy there is in an electron, so we cannot tell whether its mass is zero or not.

As we have seen, the momentum of an electron or any other kind of particle is given by the expression $\mathcal{E}\mathbf{v}/c^2$, where $\mathcal{E}$ is its energy. If we take the energy of the electron to be not only its internal energy but also the energy of its electromagnetic field which moves along with it, then we may consider it as a particle having momentum given by the above expression, and we may regard the forces which its own field exerts on it as internal forces which do not modify the motion of its centre of mass. The force on it may then be taken to be $e\left(\mathbf{F}+\frac{1}{c}[\mathbf{V}\cdot\mathbf{H}]\right)$, where $\mathbf{F}$ and $\mathbf{H}$ are the external electric and magnetic fields. When the velocity of the electron changes it may emit radiation, that is, a field which does not move along with it. Any such radiation field must be regarded as belonging to the external field and not to the field of the electron which moves along with it, but provided the acceleration of the electron is small the effect of the radiation emitted will be negligible. The electron may therefore be regarded as a particle of mass $m=\frac{m_0}{\sqrt{1-v^2/c^2}}$, where $m_0=\mathcal{E}_0/c^2$ and $\mathcal{E}_0$ is the energy of

the field of the electron which moves along with it together with its internal energy, both reckoned as when the electron is at rest.

It is important to note that the radius of an electron is not known. It is quite possible that the radius may be, say, 10^{-10} cm., in which case nearly all the mass would be mass of the internal energy, and the mass of the field of the electron would be negligible. In this case the complications which are sometimes introduced by supposing the internal energy to be zero would be quite unnecessary. All the experimental results on electrons agree with the view that the electron has a charge e and mass $m = \dfrac{m_0}{\sqrt{1 - v^2/c^2}}$, and speculations as to the nature of this mass beyond saying that it must be the mass of energy equal to mc^2 are not justified at present.

14. Radiation from an Accelerated Electron.

The radiation from an electron when its velocity varies may be calculated by means of the solution of the electromagnetic equations discussed above, but the calculation is complicated and the desired result may be obtained much more easily by a method due in principle to J J Thomson.

Consider an electron moving along a straight line AO with constant velocity v. At and near to O suppose the velocity changes in a short time δt to a constant velocity v' along OB. We suppose v/c and v'/c to be very small (fig. 2). While the electron is moving along AO its field moves with it. At O where v changes it will begin to excite the field corresponding to the velocity v' along OB, and at a time interval t later this new field will fill a sphere of radius ct. Outside this sphere the field will still be that due to the electron moving with velocity v along AO. The two fields will be separated by a layer of thickness $c\delta t$ containing the field excited by the electron during the short interval δt in which its velocity changed from v to v'. This layer moves out with the velocity c and it contains the wave produced by the change from v to v'.

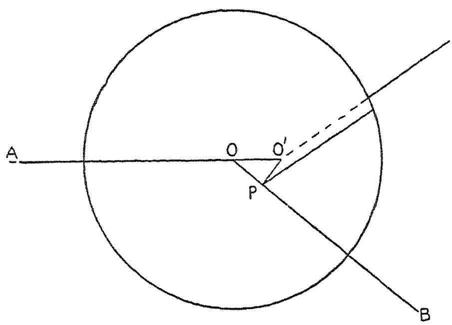


Fig 2

The lines of force in the field outside the sphere of radius ct will radiate from a point O' on AO produced such that $OO' = vt$, and the lines of force inside the sphere radiate from a point P on OB such that $OP = v't$. If we consider a line of force starting from P and making an angle θ with $O'P$, it will be displaced relatively to the parallel line outside the sphere by a distance $O'P \sin\theta$, and we may suppose these two lines are joined into a single line by a part lying in the layer of thickness $c\delta t$. This requires a field component in the layer, in the plane containing $O'P$ and parallel to the surface of the sphere, equal to $\dfrac{e}{4\pi r^2}\dfrac{O'P\sin\theta}{c\delta t}$, where $r = ct$, because the radial component in the layer is equal to $e/4\pi r^2$. When t is zero the line of force considered is all outside the sphere of

radius ct, and as t increases the relative displacement $O'P \sin\theta$ increases proportionally to t, so that it is clear that a line inside the sphere corresponds to a parallel line outside. Let $O'P = v''t$, so that the tangential field in the layer is $\frac{ev'' \sin\theta}{4\pi r c^2 \delta t}$.

The energy in the wave per unit volume is therefore

$$\left(\frac{ev'' \sin\theta}{4\pi r c^2 \delta t}\right)^2,$$

since there must be a magnetic field of equal strength to the electric field in the wave and the energy in unit volume is $\frac{1}{2}(F^2 + H^2)$. The energy in the whole wave is therefore

$$\int_0^\pi 2\pi r \sin\theta\, r\, d\theta \left(\frac{ev'' \sin\theta}{4\pi r c^2 \delta t}\right)^2 c\delta t = \frac{1}{6\pi} \frac{e^2 v''^2}{c^3 \delta t}.$$

This is the energy radiated in the time δt while the velocity changed from v to v', so that the rate of radiation during δt is

$$\frac{1}{6\pi} \frac{e^2 v''^2}{c^3 (\delta t)^2}.$$

But v'' is the vector difference between v' and v, so that $v''/\delta t$ is the acceleration of the electron during the time δt, and the rate of radiation of energy is

$$\frac{1}{6\pi} \frac{e^2 f^2}{c^3},$$

where f is the acceleration of the electron.

Electromagnetic radiation is obtained in practice from electrical oscillations produced by the discharge of a condenser through a wire. In such cases, in which enormous numbers of electrons are involved the radiation obtained agrees with that calculated by electromagnetic theory. Radiation from single electrons has not been observed, and according to the Quantum Theory the electrons in atoms do not radiate when they are moving round orbits and so have an acceleration. The success of the quantum theory makes it possible that the expression just obtained for the radiation from an electron is erroneous, and in fact that the equations of the electron theory are probably only true when the density of electricity ρ is taken to be the average density over a volume containing a large number of electrons and atomic nuclei. Another possibility is that the electricity in atoms is not really concentrated into small particles at all, but that the volume of an electron is comparable with the volume of an atom, so that the electrons in atoms overlap each other.

PROPERTIES OF MATTER IN BULK

15. The electron theory gives a fairly satisfactory explanation of the principal electrical properties of matter in bulk, such as specific inductive capacity and conductivity. The specific inductive capacity (K) of a substance may be defined as the ratio of the capacity of a condenser in which the substance fills the space between the plates to the capacity with a vacuum between the plates. For the same charges the electric field is inversely as K. The electric field F inside a material body, as for

example the insulator between the plates of a condenser, may be considered either microscopically or macroscopically. From the microscopic standpoint it is the field at a point inside the body due to all the electrons and nuclei present and it varies rapidly from point to point, being very large at points close to electrons or nuclei. From the macroscopic standpoint these variations are ignored, and we consider the field as equal to its average value over a space containing an enormous number of electrons and nuclei. Thus if $\overline{\mathbf{F}}$ denotes this average value then

$$\overline{F}_x = \frac{1}{S}\int F_x dS,$$

where $\overline{F}_x$ is the x component of the macroscopic field and F_x the x component of the microscopic field, S being the small volume over which the average is taken. The difference of potential between two points is given by

$$V_2 - V_1 = \int F_s ds,$$

where F_s is the microscopic field component along an element ds of the path between the two points. It is easy to see that $\overline{V}_2 - \overline{V}_1$ is also equal to $\int \overline{F}_s ds$, since $V_2 - V_1$ is independent of the path. In experimental work the field is taken to be $\overline{\mathbf{F}}$ and is generally estimated by measuring a potential difference and dividing this by the appropriate length. The density of charge in matter may also be regarded macroscopically and we may define the macroscopic density by

$$\bar{\rho} = \frac{1}{S}\int \rho dS.$$

When $\bar{\rho} = 0$, what is meant is that the small volume S contains as much positive as negative electricity, or if it contains n electrons each having a charge e, and n' nuclei each having a charge e', then $ne + n'e' = 0$. The electrons and nuclei in matter may be moving about so that, strictly speaking, ρ should be averaged not only over a small volume but also over a short time interval to get $\bar{\rho}$. The rapid fluctuations due to the motions of the electrons and nuclei are also ignored or averaged out from the macroscopic view-point. The microscopic equations $\operatorname{div} \mathbf{H} = 0$, $\operatorname{div} \mathbf{F} = \rho$, $\operatorname{curl} \mathbf{H} = \frac{1}{c}(\dot{\mathbf{F}} + \rho \mathbf{V})$, and $\operatorname{curl} \mathbf{F} = -\frac{1}{c}\dot{\mathbf{H}}$ of the electron theory can be easily transformed into equations between the macroscopic or average values of $\mathbf{F}$, $\mathbf{H}$, ρ, and $\rho\mathbf{V}$ given by

$$\overline{\mathbf{F}} = \frac{1}{S}\int \mathbf{F} dS, \quad \overline{\mathbf{H}} = \frac{1}{S}\int \mathbf{H} dS, \quad \bar{\rho} = \frac{1}{S}\int \rho dS, \quad \text{and } \overline{\rho\mathbf{V}} = \frac{1}{S}\int \rho\mathbf{V} dS.$$

For a non-magnetic insulating material medium which is at rest we get

$$\operatorname{div} \bar{\mathbf{H}} = 0, \quad \operatorname{div} \bar{\mathbf{F}} = \bar{\rho}, \quad \operatorname{curl} \bar{\mathbf{H}} = \frac{1}{c}(\dot{\bar{\mathbf{F}}} + \overline{\rho \mathbf{V}}), \quad \text{and} \operatorname{curl} \bar{\mathbf{F}} = -\frac{1}{c}\dot{\bar{\mathbf{H}}}.$$

16. Polarization. Electric Displacement. Equations for Non-magnetic Insulator at Rest.

The material medium is supposed to contain electrons and positive nuclei, and in the absence of an electromagnetic field the macroscopic field due to the negative charges is supposed to be equal and opposite to that due to the positive charges. When there is a field in the medium the positive charges are displaced relative to the negative charges, so that the medium acquires an electric moment or polarization $\bar{\mathbf{P}}$. The polarization $\bar{\mathbf{P}}$ is defined to be the macroscopic or average electric moment per unit volume. $\bar{\mathbf{P}}$ is a vector and we can draw lines to represent its direction. The number of lines per unit area drawn perpendicular to the lines of polarization is taken equal to $\bar{\mathbf{P}}$. If an electron having a charge e in the medium is displaced a distance ξ from its normal position an electric moment $e\xi$ will be produced, since we suppose that when all the charges are in their normal positions the average moment is zero. The total moment of a small volume S will therefore be equal to $\Sigma e\xi$, where the sign Σ indicates that the products $e\xi$ for all the positive and negative charges in S are to be summed. ξ of course is a vector, so that $\Sigma e\xi$ indicates the vector sum of all the $e\xi$'s in S. The polarization $\bar{\mathbf{P}}$ is then equal to $\frac{1}{S}\Sigma e\xi$. If we draw lines of polarization through the boundary of any small area we get a tube of polarization. Consider a short length ds of such a tube and let its cross-section be α, so that its volume is αds and its electric moment $\mathbf{P}\alpha ds$. This moment is equal to that of a positive charge $\bar{\mathbf{P}}\alpha$ and a negative charge $\bar{\mathbf{P}}\alpha$ at a distance ds apart. Thus we see that if a tube of polarization starts in the medium it must start from a charge $-\bar{\mathbf{P}}\alpha$, and if a tube ends in the medium it must end on a charge $+\bar{\mathbf{P}}\alpha$. The number of lines of polarization in the tube is $\bar{\mathbf{P}}\alpha$, so that one line of polarization starts from a unit of negative charge. We see therefore that $\operatorname{div} \bar{\mathbf{P}} = -\bar{\rho}_P$, where $\bar{\rho}_P$ denotes the average charge density due to variation of the polarization. Let $\bar{\rho}_E$ denote the average charge density due to electricity in the medium, not due to variations in the polarization. The charge $\bar{\rho}_E$ would remain when the field was reduced to zero but the charge $\bar{\rho}_P$ would disappear. The equation

$$\operatorname{div} \bar{\mathbf{F}} = \bar{\rho}$$

therefore gives, since $\bar{\rho} = \bar{\rho}_P + \bar{\rho}_E$,

$$\operatorname{div} (\bar{\mathbf{F}} + \bar{\mathbf{P}}) = \bar{\rho}_E.$$

Now consider the term $\overline{\rho \mathbf{V}}$. We have

$$\overline{\rho V_x} = \frac{1}{S}\int \rho V_x\, dS.$$

Here ρ is the charge density in dS and V_x the x component of the velocity of the electricity. Since the electricity is supposed to consist of small particles of which a very large number are in the volume S, we may replace $\int \rho V_x\, dS$ by $\Sigma e V_x$, which denotes the sum of the products eV_x for all the charges in S.

But $\Sigma e V_x = \frac{d}{dt}\Sigma e \xi_x$, where ξ_x is the x component of the displacement of a particle from its normal position. Hence, since $\overline{P}_x = \frac{1}{S}\Sigma e\xi_x$, we get

$$\dot{\overline{\mathbf{P}}} = \frac{1}{S}\Sigma e\mathbf{V}.$$

The equation $$\operatorname{curl} \overline{\mathbf{H}} = \frac{1}{c}(\dot{\overline{\mathbf{F}}} + \overline{\rho \mathbf{V}})$$

therefore becomes $$\operatorname{curl} \overline{\mathbf{H}} = \frac{1}{c}(\dot{\overline{\mathbf{F}}} + \dot{\overline{\mathbf{P}}}) = \frac{1}{c}\dot{\overline{\mathbf{D}}},$$

where $\overline{\mathbf{D}} = \overline{\mathbf{F}} + \overline{\mathbf{P}}$. $\overline{\mathbf{D}}$ is called the electric induction or the total polarization, $\overline{\mathbf{F}}$ being the polarization of the electric field and $\overline{\mathbf{P}}$ the additional polarization due to the matter present. $\overline{\mathbf{D}}$ is also sometimes called the electric displacement.

The equations for the non-magnetic insulator at rest may now be written, with the dashes omitted, in the form

$$\operatorname{div} \mathbf{H} = 0, \quad \operatorname{div} \mathbf{D} = \rho,$$
$$\operatorname{curl} \mathbf{H} = \frac{1}{c}\dot{\mathbf{D}}, \quad \operatorname{curl} \mathbf{F} = -\frac{1}{c}\dot{\mathbf{H}}.$$

Here, $\mathbf{H}$, $\mathbf{F}$, $\mathbf{D}$, and ρ now stand for the macroscopic average values, which can be determined experimentally.

17. Specific Inductive Capacity. Refractive Index.

The specific inductive capacity K of the medium may be defined as the ratio of the total polarization $\mathbf{D}$ to the field strength $\mathbf{F}$, so that $K = \mathbf{D}/\mathbf{F}$. Hence

$$\operatorname{curl} \mathbf{H} = \frac{1}{c}K\dot{\mathbf{F}}, \quad \operatorname{curl} \mathbf{F} = -\frac{1}{c}\dot{\mathbf{H}}.$$

These equations give $\operatorname{curl}\operatorname{curl} \mathbf{H} = -\Delta\mathbf{H} = \operatorname{curl}\left(\frac{1}{c}K\dot{\mathbf{F}}\right)$,

or $$\Delta\mathbf{H} - \frac{K}{c^2}\ddot{\mathbf{H}} = 0.$$

We see therefore that electromagnetic waves will travel in the medium with the velocity $c/\sqrt{K}$, so that the refractive index of the medium relative to a vacuum is equal to $\sqrt{K}$. Since $\mathbf{P} = \Sigma e\xi$, where the Σ now indicates the vector sum of the products $e\xi$ for a unit volume, we have

$$K = \frac{\mathbf{F} + \mathbf{P}}{\mathbf{F}} = 1 + \frac{\Sigma e\xi}{\mathbf{F}}.$$

The relation between $\Sigma e\xi$ and $\mathbf{F}$ may be found if we suppose that the charges in the insulator are acted on by elastic restoring forces proportional to the displacements ξ from the normal positions. The force on a charge e will be the restoring force $-\alpha\xi$, where α is a constant, the force $\mathbf{F}e$ due to the macroscopic electric field, and the force due to any field arising from the other electrons and nuclei near The actual field at a particle is the microscopic field there, and so is not necessarily equal to $\mathbf{F}$. The field at the particle may be estimated by supposing that the small space around it, in which there are no other charges, may be regarded as a small spherical cavity in the medium with the particle at its centre. For the purpose of calculating the field in such a cavity we may regard the medium as consisting of a uniform distribution of positive electricity of density ρ, and a uniform distribution of negative electricity of equal density $-\rho$, so that the total density is zero. When the medium is polarized we suppose the positive electricity displaced relative to the negative in the direction of the polarization through a distance ξ given by $\mathbf{P} = \rho\xi$. The field due to a polarized sphere at a point inside it is then equal to the field due to a positive sphere together with that due to an equal negative sphere, the distance between the centres of the two spheres being equal to ξ. The field due to a sphere of density ρ at a distance r from its centre is equal to $\frac{4}{3}\pi\rho r^3/4\pi r^2$ or $\rho r/3$, so that the resultant field due to the two spheres is equal to $-\rho\xi/3$ at any point inside the two spheres. The field due to the charges on the walls of a spherical cavity in the medium is equal and opposite to that due to the sphere removed to make the cavity, so that the total field in the cavity is equal to $\mathbf{F} + \frac{\rho\xi}{3}$ or $\mathbf{F} + \frac{\mathbf{P}}{3}$. It is sometimes stated that the field due to a sphere uniformly filled with polarized atoms all equally polarized in the same direction is equal to zero, but this statement is clearly erroneous. The demagnetizing field, for example, in a uniformly magnetized sphere is well known to be equal to $\frac{\mathbf{I}}{3}$, where $\mathbf{I}$ is the intensity of magnetization.

If then we suppose that a charged particle in the medium can be regarded as being inside a small spherical cavity, then the force on it will be equal to $e(\mathbf{F} + \mathbf{P}/3)$, due to the macroscopic field $\mathbf{F}$ and the polarization $\mathbf{P}$.

In the case of a steady field $\mathbf{F}$ we have therefore

$$e(\mathbf{F} + \mathbf{P}/3) - \alpha\xi = 0.$$

This, with $K = 1 + \frac{\Sigma e\xi}{\mathbf{F}}$ and $\mathbf{P} = \mathbf{F}(K-1)$, gives

$$\frac{K-1}{K+2} = \tfrac{1}{3}\Sigma\frac{e^2}{\alpha}.$$

If the electric field is not steady, then it is necessary to take into account the inertia of the charged particles. We may suppose that the nuclei remain fixed so that only the electrons have to be considered For a particular electron of mass m and charge e we have then (dropping the vector notation)

$$m\ddot{\xi} = -\alpha\xi + e(F + P/3).$$

If we assume for simplicity that α is the same for all the electrons, we get (since $P = \Sigma e\xi$ so that $P = \Sigma e\xi = ne\xi$)

$$nme\ddot{\xi} = -ne\alpha\xi - ne^2(F + P/3)$$

or

$$mP = -\alpha P + ne^2(F + P/3),$$

where n is the number of electrons in unit volume. Taking $F = F_0\varepsilon^{ipt}$ and $P = P_0\varepsilon^{ipt}$, we get

$$-mp^2P_0 = -\alpha P_0 + ne^2(F_0 + P_0/3),$$

so that

$$K = 1 - \frac{P}{F} = 1 + \frac{ne^2}{\alpha - \frac{ne^2}{3} - mp^2}.$$

Thus the specific inductive capacity and so the refractive index $\sqrt{K}$ vary with the frequency $p/2\pi$. In this way a satisfactory explanation of the phenomena of dispersion can be deduced from the electron theory. By introducing a frictional force on the electrons proportional to the velocity $\dot{\xi}$, absorption can also be explained. These questions are more fully discussed in works on Light.

18. Equations for a Ferromagnetic Substance at Rest.

So far we have supposed the medium to be a non-magnetic insulator at rest. In the case of a ferromagnetic substance, on the electron theory, it is supposed that the atoms contain electrons describing orbits in such a way that the atoms have a magnetic moment, or else that the atoms contain spinning electrons which have a magnetic moment. The magnetic moments of the atoms are assumed to be practically constant, and the magnetic properties are explained by the way in which the magnetic atoms are arranged. This question is discussed in the chapter on Theories of Magnetism. Here we shall consider the macroscopic electromagnetic equations for a ferromagnetic substance at rest. In the case of the non-magnetic insulator we supposed that the charges could be regarded as neutralizing each other in the absence of an electric field. The normal position of an electron can be regarded as its average position over a time long compared with its period of revolution in its orbit, and the displacement $\boldsymbol{\xi}$ as a displacement of this average position. This rather crude procedure will evidently not suffice in the case of a ferromagnetic substance, because the magnetic moments of the atoms are produced by the orbital motion. The macroscopic value of $\rho\mathbf{V}$ given by

$$\overline{\rho\mathbf{V}} = \frac{1}{S}\int \rho\mathbf{V}dS$$

does not include the orbital motion of the electrons, because the volume S is supposed to include a great many atoms, so that the orbital motions will cancel each other since they are as likely to be in one direction as in the opposite direction. The magnetic properties are due to micro-

scopic convection currents although the macroscopic convection current is zero.

If an electron is describing an orbit of area α with frequency n the average current round the orbit is ne, so that the average magnetic moment of the orbit is $n\alpha e/c$. The intensity of magnetization due to such orbits is therefore given by

$$\mathbf{I} = \Sigma n e \alpha / c,$$

where the Σ indicates the vector sum of the moments of all the orbits in unit volume.

A uniformly magnetized bar may be compared with a closely packed bundle of solenoids each of small cross-section. The magnetic field in a uniform solenoid wound with n turns of wire per unit length is equal to nC/c, where C is the current in Heaviside electrostatic units. The magnetic moment of the solenoid is equal to $Cn\alpha l/c$, where α is the cross-section of the solenoid. The magnetic moment per unit volume of the solenoid is therefore equal to Cn/c, which is equal to the field strength in the solenoid. Thus the average field strength in a magnetized substance due to the magnetization is equal to the intensity of magnetization. The total macroscopic field strength or the magnetic induction is therefore $\mathbf{B} = \mathbf{H} + \mathbf{I}$, where $\mathbf{H}$ is the field due to external sources. The induction $\mathbf{B}$ is simply the average magnetic field, so that $\mathbf{B} = \overline{\mathbf{H}} = \dfrac{1}{S}\int \mathbf{H} dS$. The current $\dot{\mathbf{D}}/c$, however, does not include the microscopic orbital currents, so that a term to represent them must be added. The macroscopic equations for a magnetic insulator at rest are therefore

$$\operatorname{div} \mathbf{B} = 0, \quad \operatorname{div} \mathbf{D} = \rho,$$

$$\operatorname{curl} \mathbf{B} = \frac{1}{c}\dot{\mathbf{D}} + \operatorname{curl} \mathbf{I}, \quad \operatorname{curl} \mathbf{F} = -\frac{1}{c}\dot{\mathbf{B}}.$$

The third equation is equivalent to curl $\mathbf{H} = \dot{\mathbf{D}}/c$, where $\mathbf{H}$ is the part of the magnetic field inside the substance not due to the local magnetization. It may be defined in the usual way as the force on a unit pole put in a long narrow cavity in the medium, the sides of the cavity being parallel to the direction of magnetization. This division of the magnetic field into two components $\mathbf{I}$ and $\mathbf{H}$ is arbitrary, but it is convenient, because of the difficulty of dealing with the orbital convection currents otherwise than by saying that they produce the magnetization $\mathbf{I}$.

We may remark here that, according to the recent new developments of the quantum theory, the charge of an electron describing an atomic orbit is regarded as distributed all round the orbit, so that the orbit is really much more nearly equivalent to a steady current of moment

$ne\alpha/c$ than it would be if the charge e were concentrated at a point.

If we assume the permeability $\mu = \mathbf{B}/\mathbf{H}$ to be a constant, and put $\mathbf{D} = K\mathbf{F}$, the equations become

$$\operatorname{div} \mathbf{H} = 0, \quad \operatorname{div} K\mathbf{F} = \rho,$$
$$\operatorname{curl} \mathbf{H} = K\dot{\mathbf{F}}/c, \quad \operatorname{curl} \mathbf{F} = -\mu\dot{\mathbf{H}}/c.$$

These equations give

$$\Delta\mathbf{H} = -\operatorname{curl}\operatorname{curl}\mathbf{H} = -\operatorname{curl}(K\dot{\mathbf{F}}/c),$$

so that
$$\Delta\mathbf{H} - \frac{\mu K}{c^2}\ddot{\mathbf{H}} = 0,$$

which shows that electromagnetic waves travel in the medium with velocity $c/\sqrt{\mu K}$, so that the refractive index is equal to $\sqrt{\mu K}$.

So far we have considered the substance to be an insulator. If it is a conductor, then the conduction current must be added to $\dot{\mathbf{D}}$ in the third equation. If the conductivity is σ, the conduction current, which of course is a convection current or stream of electrons, is equal to $\sigma\mathbf{F}$.

19. Equations for a Moving Medium.

If the material medium is not at rest, as we have hitherto supposed it to be, then it is necessary to add new terms to the electromagnetic equations to take account of the convection currents due to the motion. If $\mathbf{W}$ denotes the velocity of the medium then there is a convection current $\mathbf{W}\rho$ due to the charge per unit volume. This charge ρ, as before, is the measurable charge and does not include the charge due to the polarization. The variation of the polarization $\mathbf{P}$ gives a current $\dot{\mathbf{P}}/c$ as when the medium is at rest, but in addition the motion of the polarization with the medium produces a magnetic field. The value of this field may be calculated by regarding the polarized medium as consisting of a series of electrical double layers, each layer having a positive charge S per unit area and a negative charge $-S$ separated from the positive layer by a distance ξ. The moment of a layer per unit area is then $S\xi$, so that $\mathbf{P} = NS\xi$, where N is the number of layers per unit length. When the medium is moving with velocity $\mathbf{W}$ the double layers move with velocity $W\sin\theta$ parallel to themselves, where θ is the angle between $\mathbf{P}$ and $\mathbf{W}$. There is then a convection current $SW\sin\theta$ per unit length along the positive layers, and $-SW\sin\theta$ along the negative layers. These currents give a magnetomotive force in the layers equal to $\frac{1}{c}SW\sin\theta$ in a direction perpendicular to $\mathbf{P}$ and to $\mathbf{W}$. The average strength of this force is therefore $\frac{1}{c}N\xi SW\sin\theta$, or $\frac{1}{c}PW\sin\theta$ The force is therefore equal to the vector product of $\mathbf{P}$ and $\mathbf{W}$ into $\frac{1}{c}$, or $\frac{1}{c}[\mathbf{P}\cdot\mathbf{W}]$.

The equation
$$\operatorname{curl}\mathbf{H} = \frac{1}{c}\dot{\mathbf{D}}$$

for a non-magnetic insulator at rest becomes therefore

$$\text{curl}\,(\mathbf{H} - [\mathbf{P}\cdot\mathbf{W}]) = \frac{1}{c}(\dot{\mathbf{D}} + \rho\mathbf{W} + \mathbf{F}\sigma),$$

or
$$\text{curl}\,\mathbf{H} = \frac{1}{c}(\dot{\mathbf{D}} + \rho\mathbf{W} + \mathbf{F}\sigma + \text{curl}\,[\mathbf{P}\cdot\mathbf{W}])$$

for a non-magnetic conductor moving with velocity $\mathbf{W}$. $\dot{\mathbf{D}}$ is the displacement current, $\rho\mathbf{W}$ the convection current, $\mathbf{F}\sigma$ the conduction current, and curl $[\mathbf{P}\cdot\mathbf{W}]$ the convection current due to the motion of the polarization. The equations

$$\text{div}\,\mathbf{D} = \rho,\ \ \text{div}\,\mathbf{H} = 0,\ \text{and}\ \ \text{curl}\,\mathbf{F} = -\frac{1}{c}\dot{\mathbf{H}}$$

remain unchanged

In the case of a ferromagnetic body moving with velocity $\mathbf{W}$ we may regard the electron orbits as equivalent to magnetic doublets, and so the magnetization $\mathbf{I}$ may be considered equivalent to a series of magnetic double layers The motion of these double layers produces an electromotive force equal to $-\frac{1}{c}[\mathbf{I}\cdot\mathbf{W}]$, just as the electrical double layers give a magnetomotive force $\frac{1}{c}[\mathbf{P}\cdot\mathbf{W}]$. The equations for a magnetic moving body are therefore

$$\begin{aligned}
&\text{div}\,\mathbf{B} = 0,\ \ \text{div}\,\mathbf{D} = \rho,\\
&\text{curl}\,\mathbf{B} = \frac{1}{c}(\dot{\mathbf{D}} + \rho\mathbf{W} + \mathbf{F}\sigma + \text{curl}\,[\mathbf{P}\cdot\mathbf{W}]) + \text{curl}\,\mathbf{I},\\
&\text{curl}\,\mathbf{F} = -\frac{1}{c}(\dot{\mathbf{B}} + \text{curl}\,[\mathbf{I}\cdot\mathbf{W}]).
\end{aligned}$$

There is no magnetic convection current analogous to $\rho\mathbf{W}$ or magnetic conduction current analogous to $\mathbf{F}\sigma$. Here $\mathbf{B}$, usually called the magnetic induction, is really the macroscopic magnetic field strength in the medium.

For an uncharged non-magnetic insulator moving in a constant electric field we have $\dot{\mathbf{D}} = 0$, $\rho\mathbf{W} = 0$, and $\mathbf{F}\sigma = 0$,

so that
$$\text{curl}\,\mathbf{H} = \frac{1}{c}\,\text{curl}\,[\mathbf{P}\cdot\mathbf{W}],$$

or
$$\mathbf{H} = \frac{1}{c}[\mathbf{P}\cdot\mathbf{W}].$$

The motion of an insulator in an electric field should therefore produce a magnetic field. This effect was detected by Rontgen by spinning a circular disc made of an insulator between the plates of a charged condenser. He obtained a magnetic field equal to that calculated.

When a material medium is moving in a magnetic field there is a force on the charges in the medium equal to $\frac{1}{c}[\mathbf{W}\cdot\mathbf{H}]$ per unit charge, due to the motion of the charges in the magnetic field. The polarization $\mathbf{P}$ when the velocity and the field are constant is therefore given by $\mathbf{P} = (K-1)(\mathbf{F} + \frac{1}{c}[\mathbf{W}\cdot\mathbf{H}])$, because the polarization due to a field $\mathbf{P}$ which exerts a force $\mathbf{F}$ on a unit charge is equal to $\mathbf{F}(K-1)$. The electrical displacement $\mathbf{D} = \mathbf{P} + \mathbf{F}$ is therefore given by

$$\mathbf{D} = (K-1)\left(\mathbf{F} + \frac{1}{c}[\mathbf{W}\cdot\mathbf{H}]\right) + \mathbf{F}$$
$$= K\mathbf{F} + \frac{K-1}{c}[\mathbf{W}\cdot\mathbf{H}].$$

If $\mathbf{F}$ is zero, then
$$\mathbf{D} = \frac{K-1}{c}[\mathbf{W}\cdot\mathbf{H}]$$

In the absence of a magnetic field an equal displacement would be produced by a field $\mathbf{F}'$ given by
$$K\mathbf{F}' = \frac{K-1}{c}[\mathbf{W}\cdot\mathbf{H}],$$
so that
$$\mathbf{F}' = \frac{K-1}{K}\frac{[\mathbf{W}\cdot\mathbf{H}]}{c}.$$

This may be regarded as an electromotive force induced in the insulator by the magnetic field. The induced E.M.F. in a conductor moving in a magnetic field is equal to $\frac{1}{c}[\mathbf{W}\cdot\mathbf{H}]$, so that the induced E.M.F. in an insulator is to that in a conductor as $K-1$ is to K. This induced E.M.F. in an insulator moving in a magnetic field was measured by the writer in a hollow cylinder of ebonite rotating in a magnetic field, and was found equal to the value indicated by the electron theory.

In the case of a magnetic insulator the displacement $\mathbf{D}$ with $\mathbf{F}$ zero is
$$(K-1)\frac{[\mathbf{W}\cdot\mathbf{B}]}{c} + \frac{[\mathbf{W}\cdot\mathbf{I}]}{c}.$$

This result is obtained by putting $\mathbf{B}$ for $\mathbf{H}$ in the expression for $\mathbf{D}$ in a non-magnetic insulator when $\mathbf{F}$ is zero, and adding on the displacement corresponding to the induced E.M.F. due to the motion of the electron orbits or equivalent magnetic doublets. The field $\mathbf{F}'$ to give an equal displacement is therefore given by
$$K\mathbf{F}' = \frac{K-1}{c}[\mathbf{W}\cdot\mathbf{B}] + \frac{[\mathbf{W}\cdot\mathbf{I}]}{c}.$$

Putting $\mathbf{I} = \frac{\mu-1}{\mu}\mathbf{B}$, this gives
$$\mathbf{F}' = \frac{\mu K-1}{\mu K}\frac{[\mathbf{W}\cdot\mathbf{B}]}{c}.$$

This result was verified experimentally by M. Wilson and the writer by measuring the induced E.M.F. in a cylinder consisting of small steel balls embedded in wax, rotating in a magnetic field. This medium was used because no insulator is known for which the permeability differs appreciably from unity.

20. Metallic Conduction.

According to electron theory, the conductivity of metals is due to the presence of free electrons which move about freely between the metallic atoms. In some metals like sodium and copper there appears to be about one free electron for each metallic atom. The electrons collide with the metallic atoms, and the distance travelled between one collision and the next one is called the length of a free path. In

the absence of any electric field in the metal as many electrons, on the average, move in one direction as another, so that the current is zero; but if there is an electric field F in the metal it causes a drift of the electrons in the opposite direction to that of the field. If there are n free electrons per unit volume drifting with average velocity $\bar{v}$, the current density is $ne\bar{v}$, where e is the electronic charge.

Consider a large number of free paths all of the same length l described by electrons all moving with velocities of equal magnitude V but random direction. In fig. 3, let $OA = l$ be the radius of a sphere and OP the direction of the force Fe on the electrons. A free path starting from O along OA will be deviated by the force Fe and will end on the sphere at B. The projection of AB on OP is then equal to the displacement of the electron in the direction OP due to the field. If CB is parallel to OP, then CB is the distance the electron is displaced from its original direction, and the projection of AB on OP is equal to CB $\sin^2\theta$, where θ is the angle AOP. The mean value of CB $\sin^2\theta$ over the surface of the sphere is $\frac{2}{3}$CB and CB is equal to $\frac{1}{2}\frac{Fe}{m}\left(\frac{l}{V}\right)^2$, so the mean value of the displacement is $\frac{1}{3}\frac{Fe}{m}\left(\frac{l}{V}\right)^2$. If λ denotes the mean free path, then the mean value of l^2 is $2\lambda^2$, so the mean displacement for all free paths described with velocity V is $\frac{2}{3}Fe\lambda^2/mV^2$. An electron with velocity V describes V/λ free paths per unit time, so the velocity of drift due to the field is

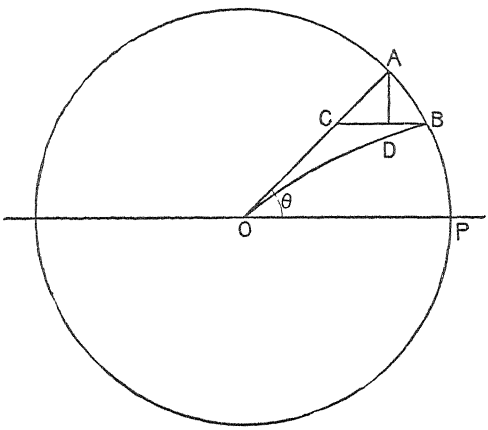


FIG. 3

$$\frac{2}{3}\frac{Fe\lambda^2}{mV^2}\frac{V}{\lambda} = \frac{2}{3}\frac{Fe\lambda}{mV}.$$

The average velocity of drift of all the electrons is then equal to $\frac{2}{3}\frac{Fe\lambda}{m}\overline{V^{-1}}$, where $\overline{V^{-1}}$ denotes the average value of $1/V$.

Before about ten years ago it was supposed that the free electrons in a metal had the same average kinetic energy and velocity distribution as the molecules of a monatomic gas. The heat capacity per mol of a monatomic gas at constant volume is equal to $\frac{3}{2}R$, where R

is the gas constant for 1 mol, which is nearly 2 calories, so the heat capacity is nearly 3 calories per mol. The heat capacities of metals are all nearly equal to 6 calories per mol at ordinary temperatures, and this value is satisfactorily explained by Debye's theory, which does not make any allowance for the heat capacity of the free electrons. If the heat capacity of the electrons is 3, the metals ought to have heat capacities equal to 9 instead of 6. It is clear, therefore, that the heat capacity of the electrons cannot be equal to 3, but must be very small. The electrons, therefore, must have kinetic energies nearly independent of the temperature, and so cannot have energy equal to that of a monatomic gas which is proportional to the absolute temperature. This difficulty was removed by the new Fermi-Dirac statistics * for an electron gas.

According to the Fermi-Dirac statistics, the number of electrons in the metal with velocities between V and $V + dV$ at low temperatures is given approximately by

$$f(V)\,dV = \frac{8\pi}{h^3}\, m^3 V^2 dV$$

for values of V between zero and V_m, but is zero for values of V greater than V_m. Also $V_m = \frac{h}{m}\left(\frac{3n}{8\pi}\right)^{1/3}$. Here m is the electronic mass, n the number of free electrons per cm.^{3}, and h Planck's constant.

If $\overline{V^{-1}}$ denotes the average value of V^{-1}, then

$$\frac{8\pi m^3}{h^3}\int_0^{V_m} \frac{V^2 dV}{V} = n\overline{V^{-1}},$$

so that $\overline{V^{-1}} = 3/2V_m$.

The average velocity of drift $\frac{2}{3}\frac{Fe\lambda}{m}\overline{V^{-1}}$ is therefore equal to $Fe\lambda/mV_m$. The current per cm.^{2} is therefore $Fne^2\lambda/mV_m$, so that the conductivity is equal to $ne^2\lambda/mV_m$.

This theory therefore explains Ohm's law and also, since the velocity distribution at low temperatures is independent of the temperature, it makes the heat capacity of the electrons zero in agreement with the facts.

It is found that the conductivity of metals is roughly inversely as the absolute temperature. To explain this we may suppose that the mean free path λ is inversely as the absolute temperature. This may be due to the amplitude of oscillation of the atoms increasing as the temperature rises.

The theory does not offer any satisfactory explanation of small second order effects such as the Hall Effect and the effect of a magnetic field on the conductivity.

* See p. 88.

21. Superconductivity.

It is found that certain metals, alloys and compounds become practically perfect conductors of electricity at very low temperatures. The following table gives the temperatures on the absolute scale below which several substances are perfect conductors.

Substance	Temperature
Lead	7·2° K
Niobium	9·2
Aluminium	1·14
Zinc	0·78
Lead sulphide	4·1
Niobium carbide	10·1
Lead-tin-bismuth alloy	8·5

Copper, silver, gold, iron, platinum and several other metals are not superconducting at 0·75° K. Solutions of metallic sodium in liquid ammonia appear to become superconducting at the temperature of liquid air.

In a magnetic field the transition temperature, at which the electrical resistance becomes zero, is lowered. For example, with lead, in a field of 200 oersteds, the temperature is 3·5° K instead of 7·2° K.

If a ring made of lead is put in a magnetic field perpendicular to the plane of the ring, cooled below 7·2° K, and then the magnetic field removed, it is found that the ring has a magnetic moment as though a current were flowing round it. This magnetic moment remains constant so long as the ring is kept below 7·2° K.

If a metal sphere is put in a uniform magnetic field and then cooled so that it becomes superconducting, the magnetic field changes so that the field inside the sphere is zero and the field outside the sphere agrees with that calculated for a sphere of zero permeability. If the sphere merely became a perfect conductor, the field would not have changed at all. It appears, therefore, that superconductors have zero permeability as well as zero electrical resistance.

In a superconductor an electron appears to move freely, so that an electric field $\mathbf{E}$ will give the electron an acceleration $\mathbf{a}$ given by $m\mathbf{a} = e\mathbf{E}$, where m is the mass of the electron and e its charge. The current density $\boldsymbol{i}$ will be equal to $ne\mathbf{V}$ where n is the number of superconducting electrons per c.c. and $\mathbf{V}$ their average velocity. We have therefore

$$\frac{d\boldsymbol{i}}{dt} = \frac{ne^2}{m}\mathbf{E}.$$

The electromagnetic equations curl $\mathbf{H} = \frac{4\pi \dot{i}}{c}$, and curl $\mathbf{E} = -\frac{\dot{\mathbf{H}}}{c}$ then give $\dot{\mathbf{H}}/c = -\frac{m}{ne^2}\,\mathrm{curl}\,\dot{i}$. Integrating this gives

$$\mathrm{curl}(i - i_0) = -\frac{ne^2}{mc}(\mathbf{H} - \mathbf{H}_0),$$

where i_0 and $\mathbf{H}_0$ are the values of i and $\mathbf{H}$ when the metal was still above its transition point. Eliminating between curl $\mathbf{H} = 4\pi i/c$ and the last equation gives $\Delta(i - i_0) = \beta^2(i - i_0)$, and similarly

$$\Delta(\mathbf{H} - \mathbf{H}_0) = \beta^2(\mathbf{H} - \mathbf{H}_0),$$

where $\beta^2 = 4\pi ne^2/mc^2$.

Consider a one-dimensional case of a superconductor bounded by a large plane area. Let x be the distance from the boundary, so that $\Delta(\mathbf{H} - \mathbf{H}_0) = \beta^2(\mathbf{H} - \mathbf{H}_0)$ becomes $\frac{d^2}{dx^2}(\mathbf{H} - \mathbf{H}_0) = \beta^2(\mathbf{H} - \mathbf{H}_0)$. The solution of this is $\mathbf{H} - \mathbf{H}_0 = Ae^{-\beta x}$ so that $\mathbf{H} = \mathbf{H}_0$ when x is large. This does not agree with the experimental results, so London assumes $\Delta\mathbf{H} = \beta^2\mathbf{H}$ and $\Delta i = \beta^2 i$. This makes $\mathbf{H} = 0$ when x is large. Also β is very large so that $\mathbf{H} = 0$, except when x is extremely small. London's theory therefore may have some truth in it.

References

1. *History of the Theories of Æther and Electricity.* E. T. Whittaker.
2. *The Theory of Electrons.* H. A. Lorentz.
3. *The Electron Theory of Matter.* O. W. Richardson.
4. "On Electromagnetic Momentum." H. A. Wilson, *Physical Review*, July, 1920.
5. *Galvanomagnetic and Thermomagnetic Effects.* L. L. Campbell.
6. "Electrodynamics of Moving Media." Swann, Tate, Bateman, Kennard (*Bulletin of the National Research Council*, Washington, D.C.)
7. "Statistical Theories of Matter, Radiation and Electricity." K. K. Darrow, *Physical Review Supplement*, July, 1929.
8. "Superconductivity." H. Grayson Smith and T. O. Wilhelm, *Reviews of Modern Physics*, Oct., 1935.

CHAPTER II

Theories of Magnetism

1. Magnetic Moment due to Electrons.

As regards their magnetic properties substances are usually classified as diamagnetic, paramagnetic, or ferromagnetic. Diamagnetic and paramagnetic bodies have constant permeabilities differing little from unity. For diamagnetic bodies the permeability is slightly less than unity, and for paramagnetic bodies slightly greater than unity. Ferromagnetic bodies have large permeabilities which vary with the magnetic field strength.

According to the electron theory an electron, with electric charge e and mass m, has angular momentum $h/4\pi$ and a magnetic moment equal to $eh/4\pi m$, where h is Planck's constant. The magnetic moment of an electron is therefore equal to $0{\cdot}93 \times 10^{-20}$ electromagnetic units. The sum of the magnetic moments of a gram-atom of electrons is therefore 5600.

An electron describing an orbit also has magnetic moment due to its motion, because a charge moving round a closed curve is equivalent to a current circuit. The nuclei of atoms also have very small magnetic moments.

The magnetic properties of substances are due to the magnetic moments of the electrons, electron orbits and nuclei in them. The magnetic properties of strongly magnetic substances like iron are almost entirely due to the magnetic moments of the electrons in the atoms of the substance.

Note that the angular momentum (often called the spin) and the magnetic moment of electrons are fundamental properties like e and m possessed equally by all electrons.

The magnetic moment of a small plane circuit carrying a current C is equal to the product of the area of the circuit and the current. An electron with charge e, describing a plane orbit of area α with frequency ν, may be regarded as equivalent to a circuit of area α and current νe, so its magnetic moment is equal to $\alpha e \nu$.* The direction of the magnetic moment is perpendicular to the plane of the orbit.

Suppose that a substance contains n electrons per unit volume, all describing plane orbits, and that α is the area of an orbit and ν the

* Cf. Chap. I, section 18. In the present chapter ordinary units are used, not Heaviside units (Chap. I, section 2), and e is in electromagnetic, not electrostatic, units.

frequency of revolution of the electron in it, then the components of the magnetic moment of the substance per unit volume or its component intensities of magnetization parallel to rectangular axes x, y, z are

$$\Sigma \alpha e \nu l, \quad \Sigma \alpha e \nu m, \quad \Sigma \alpha e \nu n$$

respectively, where l, m, n are the direction cosines of the normal to the plane of an orbit. It is understood that the sign Σ indicates the sum of the quantities like $\alpha e \nu l$ for all the n electrons. The magnetic moments of the electrons may be supposed included in this summation by making α equal to the mean area described by the electricity in the spinning electron.

2. **Theory of Diamagnetism.**

The theory of diamagnetism is closely related to that of the Zeeman Effect. The older explanation of the Zeeman Effect was in fact based on the theorem (cp. p. 415) that the equations of motion of an electron relative to fixed axes, in the absence of a magnetic field, retain very approximately the same form in the presence of a magnetic field H, provided the axes are supposed to rotate with angular velocity $\omega = He/2m$ about the direction of the field. Here e is the charge and m the mass of an electron.

We see from this that when a magnetic field is produced in a substance in which the electrons are describing orbits in fields of force, the orbits in the field will be the same relative to axes rotating about the direction of the field with velocity $\omega = He/2m$ as they were relative to fixed axes before the field was applied, provided that the application of the magnetic field does not move the fields of force.

In diamagnetic substances it is supposed that the atoms have zero magnetic moments in the absence of a magnetic field, so that the field does not tend to move them. A field H, along the x axis, therefore produces an intensity of magnetization along the x axis equal to $-\Sigma \alpha e \omega l^2/2\pi$, since the frequencies are changed by $\pm\, \omega l/2\pi$. The negative sign is correct, because an increasing magnetic field tends to induce currents which produce a field opposite to the inducing field. Hence the electrons in orbits which give a field opposite to the field H will be accelerated, and the electrons in orbits giving a field in the same direction will be retarded.

The permeability of the substance is therefore (since $\mu H = B = H + 4\pi I$) given by

$$\mu = \frac{H - \dfrac{1}{m}\Sigma \alpha e^2 H l^2}{H}$$

$$= 1 - \frac{1}{m}\Sigma \alpha e^2 l^2.$$

If we suppose that the normals to the orbits are distributed so that as many point one way as any other way, then the mean value of l^2 will be 1/3, and we have

$$\mu = 1 - \frac{e^2}{3m}\Sigma\alpha.$$

For example, a solid body having 10^{24} atoms per cubic centimetre each containing one electron orbit of area 10^{-16} sq. cm. would have a permeability given by

$$\mu = 1 - \frac{e^2}{3m}10^{24} \times 10^{-16} = 1 - 10^{-5} \text{ about.}$$

The susceptibility (κ) is defined to be $\frac{\mu - 1}{4\pi}$, so that

$$\kappa = -\frac{e^2}{12\pi m}\Sigma\alpha.$$

The value of κ/ρ, where ρ is the density of the substance, is independent of the temperature for most diamagnetic bodies, and it is independent of the field strength also. The motion of the electrons in their orbits is independent of the temperature according to Bohr's quantum theory.

3. **Paramagnetism. The Magneton.**

Paramagnetism is explained by supposing that the atoms have permanent magnetic moments. The magnetic moment of an atom is the vector sum of the moments of its electron orbits and spinning electrons. When this is zero in the absence of an external field the substance is diamagnetic, and when not zero it is paramagnetic. The effect of the magnetic field on the electron orbits which produces diamagnetism must also occur in paramagnetic substances, but it is usually too small to be appreciable compared with the paramagnetic effect of the permanent moments of the atoms.

It has been supposed that the atoms, electron orbits, spinning electrons, or other magnetic units supposed to be present in magnetic bodies always have moments which are exact multiples of an atomic unit of magnetic moment which is called a magneton. Several magnetons have been suggested, and many experimental results have been published which seem to support the view that magnetons exist. However, very accurate results are necessary to prove that a quantity is always a multiple of a definite atomic unit. This is especially the case when rather large multiples of the unit may occur. The idea of the magneton seems to be of doubtful value as far as can be judged at present.

4. **Langevin's Theory of a Paramagnetic Gas.**

A theory of a paramagnetic gas was worked out by Langevin in 1905, and this theory has been the basis of most of the subsequent theoretical work on magnetism Langevin supposed each molecule of the gas to have a fixed magnetic moment M, and the density of the gas to be so small that the mutual action of the molecules could be neglected. The molecules were supposed to contain rapidly revolving charged particles

If θ is the angle which the magnetic axis of a molecule makes with the direction of the external magnetic field H, then $M\Sigma \cos\theta$ will be the resultant magnetic moment per unit volume or intensity of magnetization I, where the sign Σ indicates the summation of the values of $\cos\theta$ for all the molecules in unit volume.

The problem is to find the value of $\Sigma \cos\theta$ as a function of H and the absolute temperature T. In Chapter V on the quantum theory the equilibrium distribution of energy among a large number of similar systems, such as atoms, each of which is supposed to be only capable of having energies E_1, E_2, E_3, . . ., is considered. It is shown that the number having energies equal to E_n is proportional to $\epsilon^{-E_n/kT}$, where k is the gas constant for one molecule and T the absolute temperature. Hence if the energy is supposed to be capable of continuous variation the number having energies between E and $E + dE$ must be proportional to $\epsilon^{-E/kT} dE$.

In a magnetic field H the potential energy of a molecule is $MH(1 - \cos\theta)$, where M is the magnetic moment and θ the angle between H and the magnetic axis. Thus we may take for dN, the number of molecules in a unit volume for which θ is between θ and $\theta + d\theta$,

$$dN = C\epsilon^{HM\cos\theta/kT} \sin\theta \, d\theta,$$

where C is a constant, since dN is proportional to $\epsilon^{-E/kT}dE$, and $E = MH(1 - \cos\theta)$ so that $dE = MH \sin\theta \, d\theta$.

To determine C we have

$$N = C\int_0^\pi \epsilon^{HM\cos\theta/kT} \sin\theta \, d\theta.$$

Let $\cos\theta = x$, so that $\sin\theta \, d\theta = - dx$,

and
$$N = C\int_{-1}^{+1} \epsilon^{HMx/kT} dx.$$

Hence
$$C = \frac{NHM}{kT\left(\epsilon^{HM/kT} - \epsilon^{-HM/kT}\right)}.$$

The magnetic moment per unit volume or intensity of magnetization is therefore given by

$$\Sigma M \cos\theta = I = \int_0^\pi \frac{NHM^2 \epsilon^{HM\cos\theta/kT} \cos\theta \sin\theta \, d\theta}{kT\left(\epsilon^{HM/kT} - \epsilon^{-HM/kT}\right)}$$

or
$$I = NM\left\{\frac{\epsilon^{HM/kT} + \epsilon^{-HM/kT}}{\epsilon^{HM/kT} - \epsilon^{-HM/kT}} - \frac{kT}{HM}\right\}.$$

Hence
$$I = NM\left\{\coth\left(\frac{HM}{kT}\right) - \frac{kT}{HM}\right\}.$$

According to this, when HM/kT is large, as at very low temperatures,
$$I = N\left\{M - \frac{kT}{H}\right\};$$

and when HM/kT is small, as is usually the case,
$$I = \frac{NM^2H}{3kT},$$

so that the susceptibility is
$$\kappa = \frac{NM^2}{3kT}.$$

It was found by Curie that the susceptibility of most paramagnetic substances is approximately inversely as the absolute temperature, in agreement with Langevin's theory when HM/kT is small.

5. Extension of Langevin's Theory to Solids and Liquids.

Langevin's theory may be supposed to apply approximately to solids and liquids as well as gases, provided they have very small susceptibilities, so that the field due to the atoms is small compared with the external field H.

The intensity of magnetization I is equal to the magnetic moment of the N molecules in unit volume so that, if σ denotes the magnetic moment of 1 gm.-molecule of any substance, then
$$\sigma = \frac{\mathcal{N}M^2H}{3kT},$$

where $\mathcal{N}$ is the number of molecules in a gram-molecule of the substance. The saturation value of σ is $\sigma_m = \mathcal{N}M$, so that putting $R = k\mathcal{N}$, as usual, we get
$$\sigma = \frac{H\sigma_m^2}{3RT}.$$

This equation enables σ_m to be calculated when σ/H has been found at any temperature T. σ/H is called the molecular susceptibility and may be denoted by χ, so that $\chi = \sigma_m{}^2/3RT$.

The product χT is nearly independent of T and H for paramagnetic substances at ordinary and higher temperatures. For example, for palladium Curie found it to be constant between 22° C. and 1370° C. However, it is found not to be constant at very low temperatures.

The following are the values of χ found for anhydrous manganese sulphate at different temperatures.

Absolute Temperature.	$\chi \times 10^6$.	$\chi T \times 10^6$
14·4	636	9,160
17·8	627	11,170
20·1	603	12,100
64·9	315	20,500
77·4	275	21,300
169·6	144	24,400
293·9	88	25,900

Thus the product χT becomes much smaller at very low temperatures.

The following are the values of χ for ferrous sulphate, $FeSO_4 + 7H_2O$.

Absolute Temperature.	$\chi \times 10^6$.	$\chi T \times 10^6$.
14·7	756	11,100
20·3	571	11,600
64·6	191	12,300
77·3	160	12,400
292·3	42·4	12,400

In this case χT is much more nearly constant.

More elaborate theories of paramagnetism than Langevin's have been worked out by Gans and others. In these theories the mutual influence of the molecules is allowed for. The molecules are regarded as producing a field called the molecular field, which may be large compared with the external field H. The theory of this molecular field will be considered under ferromagnetism.

6. Modifications based on Quantum Theory.

Modifications of Langevin's theory based on the quantum theory have also been proposed by Keesom, Gans, Weyssenhoff, Reiche, and others. The general idea underlying these theories is similar to that on which the quantum theory of specific heat is based. The average energy of a gram-atom of a solid element on the classical theory is $3RT$, which makes the atomic heat equal to $3R$ or about six calories. At low temperatures the atomic heats are less than six and become very small near the absolute zero of temperature, showing that the energy per gram-atom is less than $3RT$.

According to Debye's quantum theory of specific heats the energy per gram-atom of a solid element is given by

$$E = 9R\frac{T^4}{\psi^3}\int_0^{\psi/T}\frac{x^3\,dx}{\epsilon^x - 1},$$

where ψ is a constant depending on the nature of the substance. When T is very large this formula gives $E = 3RT$ in agreement with the classical theory.

Langevin's formula,

$$I = NM\left(\coth \frac{HM}{kT} - \frac{kT}{HM}\right),$$

with $I_m = NM$, $\sigma_m = \mathcal{N}M$, and $R = \mathcal{N}k$, gives

$$\frac{I}{I_m} = \frac{\sigma}{\sigma_m} = \coth\left(\frac{H\sigma_m}{RT}\right) - \frac{RT}{H\sigma_m}.$$

We see that σ/σ_m is a function of $H\sigma_m/RT$, so that to obtain a quantum theory of paramagnetism we may replace RT in Langevin's formula by $E/3$ as given by Debye's formula. The constant ψ appropriate for calculating specific heats will be different from that to be used in the theory of paramagnetism, because paramagnetism depends only on the rotational oscillations of the molecules.

This gives
$$\frac{\sigma}{\sigma_m} = \coth\left(\frac{3H\sigma_m}{E}\right) - \frac{E}{3H\sigma_m},$$

where
$$E = 9R\frac{T^4}{\psi^3}\int_0^{\psi/T} \frac{x^3dx}{\epsilon^x - 1}.$$

At very low temperatures E is small, so that $\coth\left(\frac{3H\sigma_m}{E}\right) = 1$ and $E = \frac{3RT^4\pi^4}{5\psi^3}$.

Hence
$$\sigma = \sigma_m - \frac{RT^4\pi^4}{5H\psi^3}.$$

If $3H\sigma_m/E$ is small, we get

$$\frac{\sigma}{\sigma_m} = \frac{H\sigma_m}{E}, \text{ so that } \chi = \frac{\sigma_m^2}{E}.$$

By means of such quantum theories it is possible to explain the variation of σ with the temperature at low temperatures. The theoretical formulæ contain constants the values of which are selected so as to make the calculated results agree as well as possible with the experimental results. Under these circumstances, of course, the calculated results agree fairly well with those observed.

7. Pauli's Theory of Paramagnetism. Ferromagnetism—Weiss's Theory.

An electron is supposed to have a magnetic moment M equal to $eh/4\pi mc$,* where e is the electronic charge, m the electronic mass,

* See p. 46

h Planck's constant, and c the velocity of light. In a magnetic field H the moment M is either along the direction of the field or in the opposite direction, so each electron has magnetic energy either $-ehH/4\pi mc$ or $+ehH/4\pi mc$. If n_+ is the number of free electrons per cm.^{3} in a metal, with magnetic moments in the direction of the magnetic field H, and n_- the number with moments in the opposite direction, then the resultant moment per cm.^{3} or intensity of magnetization of the metal is $(n_+ - n_-)eh/4\pi mc$. In the absence of a magnetic field $n_+ = n_-$, so there is no magnetization.

According to the Fermi-Dirac theory,* when $H = 0$, at low temperatures, we have $n_+ = n_- = 4\pi p_m{}^3/3h^3$, where p_m is the maximum value of the momentum of an electron. The maximum energy of the electrons $E_m = p_m{}^2/2m$, with $H = 0$, so that with a magnetic field H

$$\frac{p_{m+}{}^2}{2m} = E_m + \frac{ehH}{4\pi mc} = n_+{}^{2/3}\frac{h^2}{2m}\left(\frac{3}{4\pi}\right)^{2/3}$$

and

$$\frac{p_{m-}{}^2}{2m} = E_m - \frac{ehH}{4\pi mc} = n_-{}^{2/3}\frac{h^2}{2m}\left(\frac{3}{4\pi}\right)^{2/3},$$

so that

$$\frac{ehH}{2\pi mc} = (n_+{}^{2/3} - n_-{}^{2/3})\frac{h^2}{2m}\left(\frac{3}{4\pi}\right)^{2/3}.$$

Putting $\delta n = n_+ - n_-$ so that $n_+{}^{2/3} - n_-{}^{2/3} = \dfrac{2}{3}\dfrac{\delta n}{n_-{}^{1/3}}$, we get

$$\delta n = \frac{ehH}{2\pi mc}\frac{2m}{h^2}\left(\frac{4\pi}{3}\right)^{2/3}\frac{3n_-{}^{1/3}}{2} = \frac{2eH}{hc}\left(\frac{3n}{8\pi}\right)^{1/3}.$$

The intensity of magnetization I is therefore given by

$$I = \frac{2eH}{hc}\left(\frac{3n}{8\pi}\right)^{1/3}\frac{eh}{4\pi mc} = \frac{(3n)^{1/3}e^2}{4\pi^{4/3}mc^2}H,$$

where $n = 2n_-$ is the total number of electrons per cm.^{3}, so that the susceptibility I/H is equal to $(3n)^{1/3}e^2/4\pi^{4/3}mc^2$. This theory, which is due to Pauli, gives values for the susceptibility of the alkali metals roughly equal to those observed, when n is taken equal to the number of atoms per cm.^{3}.

In Langevin's theory of paramagnetism the magnetic molecules are supposed to oscillate in the external field H without any mutual action. Langevin's theory has been modified by Weiss, so as to be applicable to ferromagnetic bodies, by supposing that in such bodies there is a

* See p. 88

"molecular field" due to the mutual action of the magnetic molecules. This molecular field is assumed to be proportional to the intensity of magnetization I and directed parallel to it. Hence

$$H_m = \beta I,$$

where H_m denotes the molecular field strength and β is a constant depending on the nature of the substance. Weiss supposes that the magnetic molecules are arranged in sets or groups which may be small crystals of which magnetic metals like iron are composed, and that these groups are usually magnetized as strongly as possible even in an apparently unmagnetized piece of the metal.

In an unmagnetized bar the magnetic axes of the groups are supposed directed at random, so that the total magnetic moment of any part of the bar containing a large number of groups is zero. An external field tends to make the magnetic axes of all the groups point along the direction of the external field, and when this has been accomplished the metal is magnetized to saturation.

Weiss supposes that Langevin's equation for a paramagnetic gas will apply to the groups of molecules in ferromagnetic bodies if H is replaced by the vector sum of H and βI, so that

$$I = NM\left\{\coth \frac{M(H+\beta I)}{kT} - \frac{kT}{M(H+\beta I)}\right\}.$$

The intensity of magnetization I' of a group when H is small compared with βI is therefore given by

$$I' = NM\left\{\coth \frac{\beta M I'}{kT} - \frac{kT}{\beta M I'}\right\}.$$

This equation gives I' as a function of the absolute temperature T. To solve it, let $\frac{\beta M I'}{kT} = x$ and plot curves showing $y = \coth x - \frac{1}{x}$ and $y = \frac{I'}{NM} = \frac{kT}{\beta N M^2} x$ as functions of x. Let x' be the value of x at which these curves intersect (fig. 1), so that

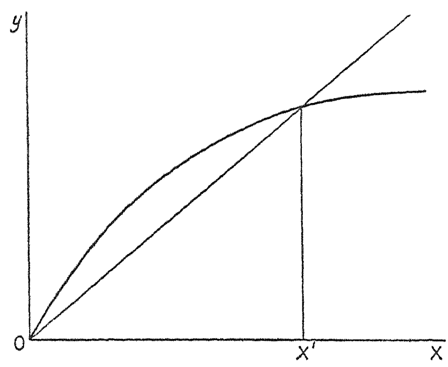


Fig. 1

$$I' = \frac{kTx'}{\beta M}$$

gives the intensity of magnetization of a group when the external field is negligible. If σ' denotes the magnetic

moment per gram-molecule for this case, then $\sigma' = I' \frac{\mathcal{N}}{N}$, where $\mathcal{N}$ is the number of molecules in a gram-molecule, so that since $\mathcal{N} k = R$,

$$x' = \frac{\beta M N \sigma'}{RT}.$$

Now if σ_m denotes the saturation value of σ, i.e. the value of the magnetic moment per gram-molecule when H is very great, then

$$\sigma_m = MN\frac{m}{D},$$

where m is the molecular weight of the magnetic molecules and D the density of the substance, so that

$$x' = \frac{\beta \sigma' \sigma_m D}{RTm}.$$

Hence
$$\sigma' = \frac{mRTx'}{\beta D \sigma_m}.$$

As the temperature is raised the slope of the line $y = \frac{kT}{\beta N M^2} x$ increases, so that the value of x' diminishes until a temperature is reached at which x' and therefore σ' become zero. This is the critical temperature at which the ferromagnetic properties disappear. When $x' = 0$, the curves $y = \coth x - \frac{1}{x}$ and $y = \frac{kT}{\beta N M^2} x$ touch each other at the origin, so that dy/dx has the same value for both curves.

When x is very small the equation

$$y = \coth x - \frac{1}{x}$$

reduces to $y = \frac{x}{3}$, so that $\frac{dy}{dx} = \frac{1}{3}$, and so at the critical temperature T_c we have

$$\frac{kT_c}{\beta N M^2} = \tfrac{1}{3},$$

or
$$T_c = \tfrac{1}{3}\frac{\beta N M^2}{k} = \frac{\beta D \sigma_m^2}{3Rm}.$$

With
$$\sigma' = \frac{mRTx'}{\beta D \sigma_m}$$

this gives
$$\frac{T}{T_c} = \frac{3\sigma'}{x'\sigma_m}.$$

Now let
$$\frac{T}{T_c} = \theta, \quad \text{and} \quad \frac{\sigma'}{\sigma_m} = \phi,$$

so that $$\theta = \frac{3\phi}{x'}.$$

We have also $$\phi = \frac{\sigma'}{\sigma_m} = \coth x' - \frac{1}{x'}.$$

The equations $$\theta = \frac{3\phi}{x'} \quad \text{and} \quad \phi = \coth x' - \frac{1}{x'}$$

give $$\phi = \coth \frac{3\phi}{\theta} - \frac{\theta}{3\phi}.$$

This equation contains nothing depending on the nature of the particular substance considered, so that if the temperature is expressed as a fraction of the critical temperature, and the intensity of magnetization of the molecular groups, in zero field, as a fraction of its saturation value, then the relation between these quantities so expressed should be the same for all substances.

Now it is found that the constant β is very large, so that the molecular field is usually very large compared with the external field H. The intensity of magnetization of the molecular groups is therefore practically uninfluenced by the external field. It is therefore equal to σ' whatever may be the value of the external field, except near the critical temperature, when σ' becomes very small. The external field therefore merely acts by lining up the molecules in the groups so that they all point along the direction of the external field. According to this the saturation intensity of magnetization observed in strong external fields is not σ_m but σ'. Thus measurements of σ in strong fields give σ', so that it is possible to test the equation $\phi = \coth \frac{3\phi}{\theta} - \frac{\theta}{3\phi}$ experimentally.

The curve in fig 2 shows the relation between θ and ϕ given by this equation, and the points marked represent the experimental results found by Weiss for magnetite in a field of 8300 gauss. The agreement is good, but in the cases of iron, nickel, and cobalt the differences between the theory and the experimental results are greater than with magnetite. Thus the theory gives a fairly satisfactory explanation of the variation of the saturation intensity of magnetization of ferromagnetic substances with the temperature.

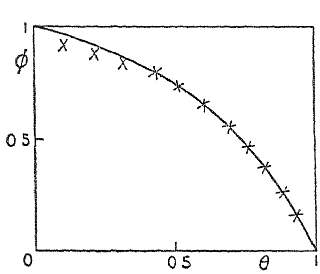


Fig 2

The value of the constant β has been found by making measurements of the intensity of magnetization near the critical temperature, where βI is not too large compared with H for the effects due to H to be

appreciable. Near the critical temperature $\frac{M(H+\beta I)}{kT}$ is small, so that we have

$$I = \frac{NM^2(H+\beta I)}{3kT}.$$

With $\quad T_c = \frac{1}{3}\frac{\beta NM^2}{k},\quad$ this gives $\quad \frac{I\beta}{H} = \frac{T_c}{T-T_c}.$

It was found by Curie that near the critical temperature the susceptibility $\kappa = I/H$ for ferromagnetic substances is inversely proportional to $T - T_c$ in agreement with this equation, so that β can be calculated from such measurements of κ. In this way it is found that β has the following values;

	β.	$\beta I'$.
Iron	3,850	6,560,000
Nickel	12,700	6,350,000
Magnetite ..	33,200	14,300,000
Cobalt	6,180	8,870,000

The values given under $\beta I'$ are the values of the molecular field in the groups of molecules at the ordinary temperature. These values are much larger than the external fields which have been used, which are seldom greater than 30,000 gauss, so that it is clear that the external fields can make no appreciable difference to I' at ordinary temperatures.

The nature of the molecular field was not known when Weiss developed his theory, but in 1928 Heisenberg showed that it could be explained by the quantum mechanical forces of exchange acting between the electrons in neighbouring atoms.

The existence of Weiss's molecular groups or domains has been confirmed experimentally. Very small colloidal particles of magnetic Fe_2O_3 are deposited on the surface of an unmagnetized piece of ferromagnetic material. The particles arrange themselves on the surface in a pattern of narrow lines about 0·01 mm. apart. At a boundary between two domains magnetized in different directions, there is a magnetic pole which attracts the magnetic particles. If the material is magnetized, the pattern changes and becomes more distinct. With a single crystal of cobalt the pattern on a surface parallel to the crystal axis consists of lines parallel to the axis, but on a surface perpendicular to the axis, a lace-like hexagonal pattern is obtained. The domains in the cobalt crystal are long thin rods parallel to the crystal axis. The domains in iron have volumes about 10^{-6} mm.^{3} or less. The magnetic moment of a domain can be changed in direction by a strong magnetic field. In weak fields the domains with moments nearly in

the direction of the field grow at the expense of adjacent domains with moments in directions not nearly in the direction of the field.

When the moment of a domain changes suddenly, the change can be detected by the induced current in a coil of wire round the magnet. If the magnet is put in a magnetic field and the field slowly increased, a telephone connected to the coil gives a click when the moment of a domain changes, so that a rapid series of clicks is produced. This is called the Barkhausen effect after its discoverer.

8. Weiss's Explanation of Hysteresis.

To explain the observed relations between the intensity of magnetization and the external field, and in particular the phenomena of hysteresis, Weiss suggested that the molecular groups which are supposed to be magnetized in random directions with intensity I' have their directions of magnetization reversed by a comparatively weak field in a direction opposite to the direction of magnetization. Let H_c be the field strength required to reverse I' in a group, and let θ be the angle between the magnetic axis of a group and the external field H. Then we suppose all directions of the magnetic axes of the groups equally probable, so that half the groups will not be affected by the external field however strong it becomes. The other half will not be affected until $H \cos\theta$ is greater than H_c, and then the directions of their axes will be reversed.

All groups for which θ is between $\pi - \bar{\theta}$ and π will be reversed, where $H \cos\bar{\theta} = H_c$, and the resultant intensity of magnetization will be twice that due to these groups. Hence

$$I = -I'\int_{\pi-\bar{\theta}}^{\pi} \sin\theta \cos\theta \, d\theta = \frac{I'}{2}\sin^2\bar{\theta},$$

$$\text{so that } I = \frac{I'}{2}\left\{1 - \left(\frac{H_c}{H}\right)^2\right\}.$$

If H is increased from 0 up to a certain value and then diminished, I will remain constant at the value given by this equation as H is diminished, until $H = -H_c$, when it will begin to diminish with H. If the maximum value of H is much greater than H_c the maximum value of I will be $I'/2$, and when H is diminished and reversed then I, when $-H$ is greater than H_c, will be given by

$$I = -\frac{I'}{2}\left\{1 - 2\left(\frac{H_c}{H}\right)^2\right\}.$$

In this case I will be zero when $H = -\sqrt{2}H_c$. In this way an hysteresis curve is obtained rather like the curves actually obtained with iron. Weiss also supposed that the molecular groups were magnetized slightly by the field H in directions perpendicular to their axes, and so obtained theoretical curves rather more like those found experimentally. However, it is clear that this theory of hysteresis is inadequate.

9. Other Theories of Ferromagnetism.

A theory of ferromagnetism similar to that of Weiss, but more elaborate, has been worked out by Gans. He assumes that in addition to the external field H and the molecular field βI there is a third field $\frac{4}{3}\pi I$. Gans supposes that the molecular field assumes different directions and that all directions are equally probable. This means that the molecular groups are continually changing the

directions of their axes. The results of Gans's theory are very similar to those of the theory of Weiss.

A theory of the relation between H and I has been developed by Honda and Okubo. This theory is an elaboration of the classical theory of Ewing. They consider first a Ewing model consisting of nine magnets. These magnets are supposed placed on a square, one at each corner, one in the middle of each side, and one in the middle of the square. The magnets are supposed free to turn in any direction except in so far as they are influenced by the external field and their own fields.

Honda and Okubo worked out the relation between the moment of such a group and the external field, and they then went on to consider the case of a large number of such groups orientated at random, which may be supposed to represent a ferromagnetic substance. They show that the relation between I and H for such a set of groups is very similar to that observed in soft iron. This agrees with the experimental results obtained long ago by Ewing, who showed that the magnetic properties of iron could be imitated by means of a large number of freely suspended small magnets equally spaced over a plane area.

Ewing has recently developed a modification of his original theory. A group of small magnets equally spaced on a lattice has too much stability to represent the behaviour of soft iron in weak fields. When the magnets are lined up so that each north pole is opposite the south pole of the next magnet it requires a very strong field to reverse the direction of the group. Ewing therefore suggests a model of a ferromagnetic atom, which consists of a small magnet free to rotate in any direction surrounded by eight fixed magnets with their centres on the corners of a cube and their axes all directed towards the small magnet at the centre of the cube. He shows that this arrangement gives the movable magnet the desired small stability, since if its south pole is near one of the fixed north poles then its north pole must be near another fixed north pole. It is then very easily deflected from one such position to another. In a group of such atoms the mutual action of the movable magnets will be added to the action of the fixed magnets on the movable ones. Ewing considers that such a model may resemble an iron atom, and that it gives a better account of the relations between field and magnetization than his original model of equally spaced freely rotatable magnets.

10. Magnetic Properties of Crystals.

The magnetic properties of several ferromagnetic crystals have been carefully investigated by Weiss and others. The case of "normal" pyrrhotite, FeS, from Brazil, will be considered here. This crystal is much more easily magnetized in one direction than in any other.

Taking this direction as the x axis, then in the yz plane there is a direction, which will be taken as the y axis, in which the crystal is much more easily magnetized than in the perpendicular direction. A field of about 15 gauss along the x axis is sufficient to reverse the direction of magnetization, and a field about double this produces saturation. The hysteresis loop for fields along the x axis resembles that obtained with very pure iron. Along the y axis a field of about 730 gauss is required to reverse the magnetization and about 12,000 gauss to produce approximate saturation. The crystal cannot be magnetized to any extent along the z axis.

The intensity of magnetization due to a field of 12,000 gauss or more in the xy plane, or the magnetic plane as it is called, is the same in all directions and is in the direction of the field. It is equal to 47.

With weaker fields the intensity is 47 when the field is along the x axis but is less along y, and for intermediate directions has intermediate values with the direction of the magnetization nearer to Ox than the direction of the field.

If H_x, H_y, H_z are the components of the external field and I_x, I_y, I_z those of the intensity of magnetization, then Weiss supposes that the total field, or resultant of the external field and the molecular field, has components $H_x + \beta_x I_x$, $H_y + \beta_y I_y$, $H_z + \beta_z I_z$, where β_x, β_y, β_z are constants which are unequal. Weiss assumes that the resultant intensity of magnetization is in the same direction as the resultant field, so that

$$\frac{H_x + \beta_x I_x}{I_x} = \frac{H_y + \beta_y I_y}{I_y} = \frac{H_z + \beta_z I_z}{I_z} = \alpha.$$

The quantity α is nearly constant for weak fields. We have therefore

$$I_x = \frac{H_x}{\alpha - \beta_x},\quad I_y = \frac{H_y}{\alpha - \beta_y},\quad I_z = \frac{H_z}{\alpha - \beta_z}.$$

If H lies in the xy plane and makes an angle θ with the x axis, then

$$\frac{H\cos\theta + \beta_x I\cos\phi}{I\cos\phi} = \frac{H\sin\theta + \beta_y I\sin\phi}{I\sin\phi},$$

where ϕ is the angle between I and the x axis. Hence

$$I = \frac{H\sin(\theta - \phi)}{(\beta_x - \beta_y)\sin\phi\cos\phi}.$$

The assumption that the intensity I is along the direction of the resultant field is equivalent to supposing that the axes of the molecular magnets are symmetrically arranged about this direction.

A similar expression to the above may easily be obtained for fields in the xz plane. It is found that the experimental results agree approximately with these formulæ. The value of $\beta_x - \beta_y$ is found to be 153 and that of $\beta_x - \beta_z$ is 3200. The saturation value of I is 47. The assumption of the existence of a molecular field enormously greater than the intensity of magnetization is difficult to justify, but it appears to give results in agreement with the facts. This field of course may represent forces on the molecules of other than magnetic origin.

The magnetic properties of magnetite, Fe_3O_4, hematite, Fe_2O_3, and of iron crystals have also been studied. They are somewhat more complicated than those of pyrrhotite.

11. Magnetization and Rotation.

Since the magnetic moment of magnetic atoms or molecules is supposed to be due to electrons describing orbits in them, or possibly

to spinning electrons, it is clear that a magnetic molecule must have a moment of momentum about its magnetic axis. It was shown in section 3 that the angular momentum due to free electrons is equal per unit volume to $2I\frac{m}{e}$, where I is the intensity of magnetization, m the mass, and e the charge of an electron. It can easily be shown that this result is true for magnetization due to electrons describing orbits of any kind. According to this, when a bar is magnetized it acquires internal angular momentum $2I\frac{m}{e}$ per unit volume. Since the total angular momentum of the bar must remain constant provided there is no couple acting on it, we should expect the bar to be set rotating about its magnetic axis with angular momentum equal and opposite to the internal angular momentum. This was pointed out by O. W. Richardson in 1914. The effect has since been detected experimentally by Einstein and Haas, and has been carefully measured by several observers. It is found that the observed effect is equal to $I\frac{m}{e}$ instead of $2I\frac{m}{e}$ as predicted by Richardson. The converse of Richardson's effect was discovered by Barnett in 1915. He showed that when a bar of iron is set rotating about its axis it becomes slightly magnetized.

12. **Measurement of Magnetic Moment of Atoms and Atomic Nuclei.**

Stern and Gerlach have introduced a new method of measuring the magnetic properties of atoms, and this method has already given very interesting results. Vapour of the substance to be investigated is allowed to pass through a narrow slit into a vacuum. The molecules move along straight lines in the vacuum, so forming a diverging beam. By means of a second slit parallel to the first a narrow nearly parallel beam is obtained. This beam is passed between the poles of a magnet which gives a strong field perpendicular to the plane containing the slits. The field strength is made to vary as rapidly as possible along the direction of the field. After passing through this field the molecules fall on a screen on which they are condensed and make a visible mark. In this way the deflection of the molecules by the magnetic field can be measured.

Take the x axis along the original direction of the stream of molecules from the second slit and the y axis along the direction of the magnetic field. If M denotes the magnetic moment of a molecule then the force on it due to the field is $M_y\frac{\partial H}{\partial y}$, where M_y denotes the y component of M. The deflection due to the field is therefore given by

$$y = \tfrac{1}{2}\frac{M_y}{A}\,\frac{\partial H}{\partial y}\,\frac{l^2}{v^2},$$

where M is the mass of the molecule, v its velocity, and l the distance from the second slit to the screen.

The average value of v^2 can be calculated from the temperature of the vapour, so that M_H can be deduced from the deflection.

In this way it was found that the atoms of copper, silver, and gold have magnetic moments of about 9×10^{-21}. Some of the atoms were deflected in the direction of the magnetic field as though they had moments $+9 \times 10^{-21}$, and others were deflected equally in the opposite direction as though they had moments -9×10^{-21}.

According to Bohr's quantum theory, the plane of the orbit of the outer electron must be perpendicular to the magnetic field and the electron may revolve round the orbit in either direction. The angular momentum of this electron according to the quantum theory is $h/2\pi$, where h is Planck's constant. Hence for a circular orbit of radius a we have $mva = \pm h/2\pi$, where v is the orbital velocity. The magnetic moment is the frequency $v/2\pi a$ multiplied by the electronic charge e and the area of the orbit, so that $M = \dfrac{v}{2\pi a} e\pi a^2 = \dfrac{vea}{2}$. But $va = \pm h/2\pi m$, so that $M = \pm \dfrac{he}{4\pi m}$, according to the quantum theory. We have $h = 6{\cdot}55 \times 10^{-27}$ and $e/m = 1{\cdot}77 \times 10^7$, so that $M = 9{\cdot}21 \times 10^{-21}$, which agrees well with the values found experimentally. The magnetic moment $he/4\pi m$ is sometimes called the moment of a Bohr magneton.

Atoms such as Zn, Cd, Hg, Sn, Pb, have zero magnetic moments. These atoms have two outer electrons with equal but opposite moments. Many atoms have magnetic moments which are not integral multiples of the Bohr magneton.

The magnetic moments of atomic nuclei can be accurately measured by a beautiful method invented by Rabi in 1939.

The element to be investigated is heated in a small vacuum oven so that the oven is filled with vapour of the element at a very low pressure. A narrow slit allows a stream of atoms to escape into a vacuum. A second slit about 50 cm. from the first one allows a narrow horizontal beam of atoms to pass. The beam goes into a third slit about 50 cm. from the second slit. The third slit lets the atoms into a small bulb, and the pressure in this is measured. The increase of the pressure in the bulb is proportional to the number of atoms going into it.

Between the first and second slits a magnet giving a vertical field (the strength of which varies rapidly along the vertical direction) deflects the atoms if they have a magnetic moment. Between the second and third slits a similar field deflects the beam in the opposite direction. If the magnetic moments of the atoms do not change as the atoms go through the three slits, the two fields compensate each other,

and the number of atoms getting into the third slit is not changed by the two fields.

Near to the second slit a third uniform vertical magnetic field is applied to the beam for a short distance. This uniform field does not deflect the beam; but if the atomic nuclei have angular momentum **J** and magnetic moment μ, the nuclei precess like a spinning top acted on by a couple.

The magnetic moment and angular momentum or spin are along the same direction. If this direction makes an angle θ with the magnetic field **H**, the couple on the nucleus is $\mathbf{H}\mu \sin\theta$, so that the axis of the spin revolves round the direction of **H** with angular velocity ω given by

$$\mathbf{J}\sin\theta \,.\, \omega = \mathbf{H}\mu\sin\theta \quad \text{or} \quad \mu = \mathbf{J}\omega/\mathbf{H}.$$

A weak alternating magnetic field is superposed on the uniform field **H**. This field is perpendicular to **H** and to the direction of the atomic beam. The frequency n of the alternating field is kept constant and **H** is slowly varied. When $2\pi n = \omega$, the alternating field disturbs the precession and causes the directions of the magnetic moments of the atoms to change. The atoms then do not get to the third slit because the two magnetic deflections are no longer equal and opposite. In this way ω can be accurately measured and so μ found when **J** is known.

The nuclear spin **J** can be found by spectroscopic methods so that the magnetic moment μ can be determined.

The following table gives the values of μ and **J** for several nuclei. μ is in nuclear magnetons. A nuclear magneton is equal to a Bohr magneton divided by 1840, or the ratio of the mass of the proton to that of the electron.

Element	**J**	μ
Hydrogen	1/2	2·785
Neutron	1/2	−1·935
Deuterium	1	0 855
Lithium Li^7	3/2	3·25
Aluminium	5/2	3·628
Silver Ag^{107}	1/2	−0·10
Ag^{109}	1/2	−0·19

The spin **J** is given in units equal to $h/2\pi$.

REFERENCES

1. "Magnetism." R. M. Bozorth, *Reviews of Modern Physics*, Vol. 19, No. 1, Jan., 1947.
2. *Magnetism and Atomic Structure*. E. C. Stoner.
3. *Modern Magnetism*. L. F. Bates.

CHAPTER III

Thermionics

1. Experimental Phenomena.

When a body charged with electricity is heated to a sufficiently high temperature the electricity leaks away and the body becomes discharged. The branch of physics dealing with such phenomena was called *thermionics* by O. W. Richardson, to whom our knowledge of it is largely due.

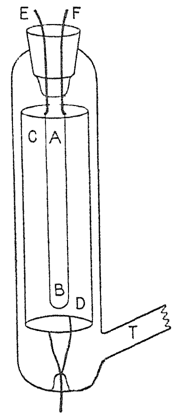


Fig 1

A simple form of apparatus for studying thermionics is shown in fig. 1. A loop of wire AB of tungsten, platinum, or any other metal having a high melting-point is supported by two platinum wires EF sealed into a glass bulb as shown. The loop is surrounded by a metal cylinder CD supported by wires sealed through the glass at the lower end of the bulb. A tube T connects the bulb to apparatus for producing a good vacuum in it.

The loop can be heated by passing a current through it, and its temperature can be found from its resistance or by means of an optical pyrometer.

The cylinder is connected through a galvanometer to one terminal of a battery by means of which any desired potential difference can be maintained between the loop and the cylinder. The other terminal of the battery is connected to one of the wires E and F. If a good vacuum is maintained in the bulb, that is if the gas pressure in it is kept below say 10^{-4} of a millimetre of mercury, then on raising the temperature of the loop it is found that a current is indicated by the galvanometer when the loop is negatively charged, but that there is no current when it is positively charged.

Negative electricity escapes from the wire to the cylinder when the wire is hot enough, but not positive electricity. The galvanometer indicates a negative current flowing into the cylinder when the cylinder is at a higher potential than the wire.

The current obtained depends on the temperature and on the

potential difference. If the temperature is kept constant, and the potential difference gradually increased from zero, then the current at first increases rapidly with the potential difference but soon attains a constant maximum value. This maximum current is called the *saturation current.* The potential required to produce the saturation current is greater when the saturation current is large than when it is small. When the temperature is raised the saturation current increases rapidly with it.

2. O. W. Richardson's Theory of the Thermionic Current.

The negative electricity which escapes from hot metals in a good vacuum consists of negative electrons. That this is the case was first shown by J. J. Thomson, who measured the ratio of the charge e to the mass m of the escaping electricity, and found it equal to about 10^7 electromagnetic units per gram, which is about the value of e/m for negative electrons.

The electrical conductivity of metals is supposed to be due to the presence in them of negative electrons which are free to move about inside the metal. It is natural to suppose that the electrons which escape from the metal at high temperatures are some of these free electrons.

The free electrons were at one time supposed to have the same average kinetic energy as the molecules of a gas at the same temperature, and a theory of the negative thermionic current based on this supposition was worked out by O. W. Richardson.

This classical theory is not now regarded as correct, but a brief account will be given of it here because of its historical interest. The more recent theories based on thermodynamics and the quantum theory will then be discussed.

We suppose that a metal contains n free electrons per unit volume and that these electrons move about inside the metal like the molecules of a gas, colliding with each other and with the metallic atoms. The average kinetic energy of the free electrons we suppose is equal to αT, where T is the absolute temperature of the metal and α is a constant.

Take axes x, y, z in the metal and let u, v, w be the components of the velocity V of an electron parallel to these axes. Then we suppose that the distribution of the velocity components among the electrons is given by Maxwell's law for gas molecules, that is, if dn denotes the number in unit volume having velocity components u between u and $u + du$, then

$$dn = nA\varepsilon^{-qu^2}du,$$

where A and q are constants at any given temperature.

We have $$\int dn = \int_{-\infty}^{+\infty} nA\varepsilon^{-qu^2}du = n,$$

and $$3\int \tfrac{1}{2}mu^2 dn = \int_{-\infty}^{+\infty} nA\tfrac{3}{2}mu^2\varepsilon^{-qu^2}du = n\alpha T.$$

But $$\int_{-\infty}^{+\infty} \epsilon^{-qu^2} du = \sqrt{\frac{\pi}{q}}, \text{ and } \int_{-\infty}^{+\infty} u^2 \epsilon^{-qu^2} du = \frac{\sqrt{\pi}}{2q^{3/2}},$$

so that $$A\sqrt{\frac{\pi}{q}} = 1, \text{ and } \frac{3mA\sqrt{\pi}}{4q^{3/2}} = \alpha T,$$

which give $$A = \sqrt{\frac{3m}{4\pi\alpha T}}, \text{ and } q = \frac{3m}{4\alpha T}.$$

For gas at pressure p containing n molecules per unit volume we have $p = \frac{1}{3}nm\overline{V^2}$, where m is the mass of one molecule and $\overline{V^2}$ the average value of the square of the velocities of the molecules.

3. The Thermionic Work Function.

Richardson assumed that before an electron can escape from the metal it has to do a definite amount of work at the surface layer. Let us denote this work by ϕe, where e is the charge on one electron, so that ϕ is the potential difference corresponding to the work in question. According to this, if $\frac{1}{2}mV_1^2$ is the kinetic energy of an electron in the metal before escaping, and $\frac{1}{2}mV_2^2$ that after escaping, then
$$\tfrac{1}{2}m(V_1^2 - V_2^2) = \phi e.$$

The number of electrons which enter the surface layer per unit area in unit time is
$$\int u\,dn = \int_0^{\infty} nA\epsilon^{-qu^2} u\,du,$$

where u is the velocity component towards the surface, which we take to be perpendicular to the x axis. Of these only those for which $\frac{1}{2}mu^2$ is greater than ϕe can escape. The number escaping per unit area in unit time is therefore
$$\int_{\sqrt{\frac{2\phi e}{m}}}^{\infty} nA\epsilon^{-qu^2} u\,du = \frac{n\epsilon^{-2\phi eq/m}}{2\sqrt{\pi q}},$$

since $A = \sqrt{\dfrac{q}{\pi}}$.

The thermionic current density i is therefore given by
$$i = \frac{ne\epsilon^{-2\phi eq/m}}{2\sqrt{\pi q}}.$$

Substituting $3m/4\alpha T$ for q, this becomes
$$i = ne\sqrt{\frac{\alpha T}{3\pi m}}\,\epsilon^{-\frac{3}{2}\frac{\phi e}{\alpha T}}.$$

If we put $$a = ne\sqrt{\frac{\alpha}{3\pi m}} \text{ and } b = \frac{3}{2}\frac{\phi e}{\alpha},$$

we get $$i = aT^{1/2}\epsilon^{-b/T}.$$

It is found that the saturation current density varies with the temperature approximately in accordance with this equation. The factor $\varepsilon^{-b/T}$ increases very rapidly with T, so that the effect of the factor $T^{1/2}$ is scarcely appreciable. However, recent very exact measurements of i seem to show that the power of T should be about 2 instead of $\frac{1}{2}$.

The following table gives some values of the thermionic work function φ expressed in volts.

Tungsten .. .	4.52
Platinum .. .	4.4
Iron	3.7
Aluminium .	3.0
Sodium . ..	1.8

The work function φ is closely related to contact potential difference. Consider a condenser consisting of two parallel plates made of different metals, and let the two plates be connected together by a wire of the same metal as one of the plates is made of.

Suppose an electron is taken from a point inside one of the plates through the wire into the other plate and then across the space between the plates back to its original position. The total work required to take the electron around this path with negligible velocity must be zero. If we neglect the small potential difference at the junction of the two metals, the work required is

$$e(\varphi_1 - \varphi_2) + e(V_2 - V_1) = 0,$$

where V_2 is the potential just outside one metal and V_1 that just outside the other. But $V_2 - V_1$ is the contact potential difference, so that the difference between the thermionic work functions for two metals is nearly equal to the contact potential difference between them. Cf. Chap. IV, section 2.

4. Objections to Classical Theory.

When light falls on the surface of a metal it may cause electrons to escape. This is the photoelectric effect (see Chap. IV). The maximum kinetic energy of the electrons which escape is given by Einstein's equation,

$$\frac{1}{2}mv^2 = h\nu - w,$$

where ν is the frequency of the light, h is Planck's constant, and w a constant depending on the nature of the metal. $h\nu$ is supposed to be the energy of one quantum of the light of frequency ν, so that if an electron in the metal absorbs a quantum then it gets energy $h\nu$.

It is found that w is equal to ϕe, the work which an electron must do in escaping through the surface layer. This seems to require that the electrons set free by the light should have no appreciable kinetic energy in the metal when they absorb the quantum of energy $h\nu$. This does not agree with Richardson's assumption that the electrons have the same average kinetic energy as gas molecules.

The specific heat of electricity in metals is also much smaller than it should be if the electrons carrying the current have the same energy as gas molecules. Also it is found that the specific heats of metals can be explained without attributing any thermal kinetic energy to the

electrons in them. For these reasons the idea that the average kinetic energy of the free electrons in metals is equal to that of gas molecules has been abandoned. It seems probable that the electrons in metals are in stationary states of the quantum theory, so that so long as they do not change from one such state to another their energy is independent of the temperature of the metal.

5. **Thermodynamical Theory of Thermionic Emission.**

If we assume that the escaping electrons form a monatomic gas or vapour which can be in equilibrium with the metal when at a definite pressure, then we can apply the thermodynamical theory of evaporation to the problem and so obtain a theory of thermionic emission without making assumptions as to the state of the electrons inside the metal A theory of this kind was first proposed by the writer and has since been developed by O. W. Richardson and others.

Let us first consider the latent heat of evaporation of one mol of electrons. Let the potential difference between a point just outside the metal and a point inside it be ϕ. Then if $\mathcal{N}$ is the number of molecules in one mol of any gas and e the charge on one electron, the work required when one mol of electrons escapes from the metal is $\mathcal{N}e\phi$.

As we have seen, the electrons in the metal have little or no kinetic energy, whereas in the gaseous state outside they have the same average kinetic energy as the molecules of a gas. For any gas (p. 50) we have $p = \frac{1}{3}mnV^2$ or $p = \frac{2}{3}naT$, since we have taken (p. 49) aT to be the average kinetic energy of the free electrons. Thus if we write $k = \frac{2}{3}a$ we have $p = nkT$, where k is the gas constant for one molecule. The kinetic energy of the molecules in one mol of a monatomic gas is therefore equal to $\frac{3}{2}\mathcal{N}kT = \frac{3}{2}RT$, where $R = \mathcal{N}k$ is the gas constant for one mol of any gas.

Also when the electrons escape as a gas at a pressure p, work pV has to be done against the external pressure p. But $pV = RT$, so that the total heat energy required to keep the temperature constant when one mol of electrons escape is

$$\mathcal{N}e\phi + \tfrac{5}{2}RT.$$

If then L denotes the heat of evaporation at the temperature T, and $L_0 = \mathcal{N}e\phi_0$ that at the absolute zero of temperature,

$$L = L_0 + \mathcal{N}e(\phi - \phi_0) + \tfrac{5}{2}RT,$$

where ϕ_0 is the value of ϕ at $T = 0$.

Contact difference of potential is found to be practically independent of the temperature, so that it is probable that $\phi - \phi_0$ is very small, and we have approximately

$$L = L_0 + \tfrac{5}{2}RT.$$

It is only allowable to regard the electron gas as a perfect gas when its pressure and its volume are very small so that the electrical forces due to the charges on the electrons can be neglected.

This introduces no difficulty, because we need not suppose that more than an infinitesimal number of electrons are outside the metal at any time. In the chapter on the quantum theory (Chap. V, section 8) it is shown that the vapour pressure of a liquid or solid which gives off a monatomic vapour is given approximately by the equation:

$$\log p = -\frac{L}{RT} + \frac{3}{2}\log T + \frac{5}{2} + \log\left\{\left(\frac{2\pi}{h^2}\right)^{3/2} k^{5/2}\right\} + \frac{3}{2}\log m,$$

where h is Planck's constant, L the heat of evaporation per mol at the temperature T, and m the mass of one molecule of the vapour. Since the electrons form a monatomic gas outside the metal we may take the pressure p at which they are in equilibrium with the metal to be given by this equation. Hence, putting $L = L_0 + \frac{5}{2}RT$, we get

$$\log p = -\frac{L_0}{RT} + \frac{5}{2}\log T + \log\left\{\left(\frac{2\pi}{h^2}\right)^{3/2} k^{5/2}\right\} + \frac{3}{2}\log m.$$

The pressure p is the very small pressure of the electron gas which is in equilibrium with the metal. The relation between p and the saturation current density i may be determined if we assume that the electrons in the gas which collide with the metal surface are all absorbed and do not rebound. This is believed to be approximately true. Assuming this, we have the result that when the gas is in equilibrium with the metal the number of electrons which escape from the metal is equal to the number which collide with it.

We suppose that the electron gas can be regarded as a perfect gas, and that the average kinetic energy of the electrons in it is the same as for any other gas at the same temperature.

The number of electrons colliding with unit area of the metal surface in unit time is therefore given by the same expression as the number entering the surface layer from the metal on Richardson's classical theory, provided that n is now taken to be the number of electrons per unit volume in the electron gas instead of inside the metal. The number striking unit area in unit time is therefore (section 2)

$$\int_0^\infty nA\epsilon^{-qu^2}u\,du = \frac{n}{2\sqrt{\pi q}}.$$

But $q = \frac{3m}{4aT}$ (p. 50), $p = nkT$, and $a = \frac{3}{2}k$ (p. 52), so that

$$\frac{n}{2\sqrt{\pi q}} = \frac{p\sqrt{2kT}}{2kT\sqrt{\pi m}} = \frac{p}{\sqrt{2\pi mkT}}.$$

The thermionic saturation current density i is therefore given by

$$i = \frac{pe}{\sqrt{2\pi mkT}},$$

where e is the charge on one electron.

Eliminating p from the equation for log p and this expression for i we get

$$i = \frac{2\pi emk^2}{h^3} T^2 \epsilon^{-L_0/RT}.$$

This theoretical expression for i is based on the thermodynamical theory of evaporation and on the quantum theory. It was first obtained by Dushman.

The values of m, k, h, and R are all known with considerable accuracy, and the only quantity in the expression for i which depends on the properties of the particular metal emitting the electrons is $L_0 = \mathcal{N}e\phi_0$.

If we put $\frac{2\pi emk^2}{h^3} = A$ and $L_0/R = b$, the equation for i becomes $i = AT^2\epsilon^{-b/T}$, which gives $b = T \log\left(\frac{AT^2}{i}\right)$. The following tables give the values of i for tungsten and platinum at different temperatures measured by Davisson, Germer, and Schlichter, and the values of b calculated from them by means of this equation.

Temperature.	Current.	b
	TUNGSTEN	
	$1 = 10^{-2}$ amp per sq cm.	
1935 5	0·0934	51,890
1986·5	0 1973	51,880
2036·0	0 3967	51,860
2077 5	0·6784	51,880
2086·5	0·7656	51,900
2102·0	0 9363	51,840
2131·5	1 362	51,840
2134·5	1 419	51,870
2158·0	1 902	51,820
2182 0	2·538	51,810
2204 0	3·269	51,820
2231·0	4 405	51,820
2235 0	4·606	51,870
2271·5	6·875	51,880
2280 0	7·394	51,920
2306·0	9·792	51,900

Temperature.	Current	b
	PLATINUM	
	$I = 10^{-5}$ amp. per sq. cm	
1211·0	0·42	49,300
1243·0	1·35	49,200
1275·0	4·17	49,100
1307·0	9·9	49,300
1339·0	27·5	49,200
1371·0	76·0	49,000
1403·0	168·0	49,100
1435·0	400·0	49,100
1467·0	785·0	49,300
1499·0	1430·0	49,500

It will be seen that the values of b are all very nearly equal, so that the equation $i = AT^2\epsilon^{-b/T}$ represents the experimental results very well indeed.

6. Fermi-Dirac Theory.

According to the Fermi-Dirac theory of an electron gas,* the number of electrons per cm.^{3} with momenta between p and $p + dp$ is equal to

$$\frac{8\pi p^2\,dp}{h^3}\,\frac{1}{\epsilon^{(E-W)/k\theta}+1},$$

where $E = p^2/2m$ and $W = \dfrac{h^2}{2m}\left(\dfrac{3N}{8\pi}\right)^{2/3}$, provided the temperature θ is not too large. N is the number of electrons per cm.^{3}. For most metals N is about 10^{23}, which makes W equivalent to about 10 electron volts.

For values of E much greater than W the number of electrons per cm.^{3} with p_x between p_x and $p_x + dp_x$ is therefore equal to

$$\frac{2dp_x}{h^3}\,\epsilon^{W/k\theta}\int_0^\infty \epsilon^{-E/k\theta}2\pi r\,dr,$$

where $r^2 = p_y^2 + p_z^2$. Putting $E = \dfrac{r^2 + p_x^2}{2m}$, this becomes

$$\frac{4\pi}{h^3}\,\epsilon^{W/k\theta}\epsilon^{-p_x^2/2mk\theta}dp_x\int_0^\infty \epsilon^{-r^2/2mk\theta}r\,dr = \frac{4\pi}{h^3}\,\epsilon^{W/k\theta}mk\theta\epsilon^{-p_x^2/2mk\theta}dp_x.$$

* See p. 88.

The number of electrons falling per second on 1 cm.[2] perpendicular to the x direction with p_x greater than p_0 is therefore

$$4\pi\epsilon^{W/k\theta}\frac{mk\theta}{h^3}\int_{p_0}^{\infty}\epsilon^{-p_x^2/2mk\theta}\frac{p_x}{m}\,dp_x.$$

If we suppose that these all escape from the metal, the thermionic current density i is given by

$$i = 4\pi\epsilon^{W/k\theta}\frac{ek\theta}{h^3}\int_{p_0}^{\infty}\epsilon^{-p_x^2/2mk\theta}p_x\,dp_x$$

or

$$i = \frac{4\pi emk^2\theta^2}{h^3}\,\epsilon^{(W-E_0)/k\theta},$$

where $E_0 = p_0^2/2m$. This result agrees with that obtained above in section (5) by means of thermodynamical theory, provided the thermionic work function ϕ is given by $\phi e = E_0 - W$. It appears, therefore, that the minimum energy E_0 which the electrons must have in order to escape is not ϕe but $\phi e + W$. ϕ is often about 4 volts, so that E_0 must be about 14 electron volts since W is equivalent to about 10 electron volts.

Davisson and Germer found that the diffraction of electron waves by metal crystals * shows that the kinetic energy of the electrons is increased by about 14 electron volts when they enter the crystal, which is what we should expect if the electrons cannot escape unless their energy inside is greater than 14 electron volts.

Richardson's classical theory of the thermionic current therefore holds good provided that the Maxwell velocity distribution is replaced by the Fermi-Dirac distribution, and the theory then agrees perfectly with the thermodynamical theory. On the Fermi-Dirac theory the energy of the electrons is practically independent of the temperature, so that the difficulty about the specific heats is satisfactorily explained.

7. Space Charge Effect.

When the potential difference used is not sufficient to produce the saturation value of the current, the current varies with the potential difference. If, for example, the current obtained is only one-half of the saturation current, then half the electrons emitted by the wire get across to the surrounding electrode and half return to the wire. The negative charge on the electrons in the space around the hot wire modifies the electric field and actually reverses it near the wire, so that the electrons coming out of the wire have to move a certain distance in a field which tends to stop them. Only those which have sufficient kinetic energy are able to get through this region in which the field is reversed, and the rest are driven back into the wire. This is called the space charge effect. To simplify the theory we will consider the case of

* See p. 95.

a large plane area emitting electrons to a large plane electrode at a distance d from the emitting plane. Let x denote the distance of a point from the emitting plane, and V the potential difference between the point and the emitting plane. Then we have

$$\frac{\partial^2 V}{\partial x^2} = -4\pi ne,$$

where ne is the density of the space charge, n being the number of electrons per unit volume and e the charge on one electron. Let the region in which the field is reversed be of thickness t. When x is greater than t all the electrons present are moving away from the emitting plane, so that if their average velocity is v we have $i = nev$.

The velocity of an electron will be given by the equation $Ve = \frac{1}{2}m(v_0^2 - v^2)$, where v_0 is the initial velocity of emission from the hot plane. The velocity components parallel to the emitting plane make no difference, so we disregard them and take v and v_0 to be perpendicular to the electrodes. In the region of reversed field V is negative so that v is less than v_0. Let the minimum potential at $x = t$ be V_1. Then if v_0 is not greater than $\sqrt{\frac{2V_1 e}{m}}$ the electron will not get across and will return to the emitting plane.

The average velocity of the electrons at $x = t$ will be the same as the average velocity of emission v_0 at $x = 0$, and we shall suppose that this is small compared with v. The average velocity will then be given approximately by $-Ve = \frac{1}{2}mv^2$. Eliminating n and v from the three equations

$$\frac{\partial^2 V}{\partial x^2} = -4\pi ne,$$

$$i = nev,$$

$$Ve = -\tfrac{1}{2}mv^2,$$

we get

$$\frac{\partial^2 V}{\partial x^2} = -2\pi i\sqrt{\frac{-2m}{Ve}}.$$

Integrating this equation we get

$$\left(\frac{\partial V}{\partial x}\right)^2 = -8\pi i\sqrt{\frac{-2mV}{e}} + \text{constant}.$$

To determine the constant, we have $\frac{\partial V}{\partial x} = 0$ at $x = t$, since V is a minimum at this point. We shall suppose that t is small, so that approximately $\frac{\partial V}{dx} = 0$ when $x = 0$ and $V = 0$, and the constant is zero.

Hence

$$\left(\frac{\partial V}{\partial x}\right)^2 = -8\pi i\sqrt{\frac{-2mV}{e}}.$$

Integrating this and putting $V = 0$ at $x = 0$,

we get $$i = -\frac{1}{9\pi}\sqrt{\frac{-2e}{m}}\,\frac{V^{3/2}}{x^2}.$$

This gives $$i = -2{\cdot}33 \times 10^{-6}\frac{V^{3/2}}{x^2},$$

where i is the current density in amperes per square centimetre, and V the potential in volts at a distance of x cm. from the hot plate. It appears that when the current is limited by the space charge it should vary as $V^{3/2}$ and inversely as x^2. This agrees approximately with the experimental results. The current is found to increase nearly as $V^{3/2}$ until it is nearly equal to the saturation current. It then remains nearly constant when V is increased further.

If instead of two parallel planes, one emitting electrons, we consider two concentric cylinders, the inside one emitting electrons, similar theoretical results are obtained.

Let b be the radius of the outer cylinder and a that of the inner one. Let V be the potential difference between the inner cylinder and a point at a distance r from its axis. The differential equation which V must satisfy is then

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} = -4\pi ne.$$

Let now i denote the current per unit length of the axis of the cylinders, so that $$i = 2\pi rnev,$$

where v is the velocity of the electrons, n the number per unit volume, and e the charge on one electron. For simplicity we shall suppose that $-Ve = \frac{1}{2}mv^2$, which means that we neglect the initial velocity of the electrons. These equations give

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} = -\frac{i}{r}\sqrt{\frac{-2m}{Ve}}.$$

We require a solution of this making $\partial V/\partial r$ zero at the radius where V is a minimum, as in the previous problem, and $V = 0$ at $r = a$. When a is small compared to b, which is usually the case in practice, then it will be approximately correct to use a solution making $V = 0$ at $r = 0$, provided this solution also makes $\partial V/\partial r$ small near the inner cylinder. If we assume $V = Ar^{2/3}$ and substitute this in the differential equation for V, we find

$$i = -\frac{2}{9}\sqrt{\frac{-2e}{m}}\,\frac{V^{3/2}}{r}.$$

This satisfies the equation, and makes $\frac{\partial V}{\partial r}$ small near the inner cylinder as well as $V = 0$ at $r = 0$, and so is an approximate solution of the problem when a/b is small.

With V in volts, r in centimetres, and i in amperes we get

$$i = -14{\cdot}65 \times 10^{-6} \frac{V^{3/2}}{b},$$

where V is now the potential difference between the outer and inner cylinders so that $r = b$. This equation is found to agree approximately with the observed currents when i is less than the saturation current and b/a is greater than about 10.

It will be observed that when the current is limited by the space charge it is independent of the temperature of the hot electrode.

Thermionic currents limited by space charge are employed in the three-electrode vacuum tubes now used so extensively as detectors and amplifiers, and for maintaining electrical oscillations.

8. Distribution of Electron Velocities.

The distribution of velocities among the electrons emitted by hot metals in a vacuum was investigated by O. W. Richardson and was found to be in agreement with Maxwell's law for the molecules of a gas.

Consider a fine straight wire emitting electrons in a vacuum and surrounded by a concentric cylindrical electrode. Let the number of electrons emitted be small so that space charge effects are negligible. If the potential difference between the wire and the cylinder is V, then when V is positive the electrons have to do work $-Ve$ in order to get across to the cylinder. Let the velocity v with which an electron is emitted be resolved into two components, v_a along the axis of the wire, and v_r along the radius of the cylinder. We suppose the radius of the hot wire very small, so that the electrons can be regarded as starting practically at the axis The component v_a does not help the electron to get across and so may be disregarded. For the electron to get across we must have

$$\tfrac{1}{2} m v_r^2 > -Ve$$

Let the number of electrons for which $\frac{1}{2} m v_r^2$ is between E and $E + dE$ be given by

$$dn = nf(E)\,dE,$$

where n is the number emitted in unit time, so that the saturation current $i_0 = ne$.

The current obtained will then be

$$i = i_0 \int_{-Ve}^{\infty} f(E)\,dE.$$

Differentiating this with respect to V, we get

$$\frac{di}{dV} = i_0 e f(-Ve).$$

Thus if we observe the current i obtained with different values of V we can determine di/dV and so get values of the function $f(E)$.

It was found experimentally that $\frac{i}{i_0} = \varepsilon^{-\alpha V}$, where α is a constant. The values found for α were nearly equal to $-e/kT$, where k is the gas constant for one molecule and T the temperature of the wire. Hence

$$\frac{1}{i_0}\frac{di}{dV} = \frac{e}{kT}\varepsilon^{Ve/kT} = ef(-Ve),$$

so that

$$f(E) = \frac{1}{kT}\varepsilon^{-E/kT}.$$

This shows that the electrons have the Maxwell velocity distribution for the velocity component v_1, for according to Maxwell's law the number of the molecules in unit volume having velocity components u between u and $u + du$ is

$$nA\varepsilon^{-qu^2}du.$$

The number of these going through a unit area perpendicular to u in unit time is therefore

$$nA\varepsilon^{-qu^2}u\,du,$$

and the total number going through is

$$nA\int_0^\infty \varepsilon^{-qu^2}u\,du = \frac{nA}{2q}.$$

Hence the fraction of those going through which have velocities between u and $u + du$ is $2q\varepsilon^{-qu^2}u\,du$

But $q = \frac{m}{2kT}$ (section 5), and $E = \frac{1}{2}mu^2$, so that the fraction is equal to $\frac{1}{kT}\varepsilon^{-E/kT}dE$, in agreement with the experimental result.

Richardson also showed that the velocity components parallel to the emitting surface are distributed in accordance with Maxwell's law

Under some circumstances hot bodies emit positively charged molecules or ions as well as electrons. This emission of positive ions will be discussed in the chapter on positive rays. It is found that some metallic oxides emit electrons at high temperatures; for example, calcium and barium oxides do this very freely. The emission of electrons by such oxides appears to vary with the temperature in much the same way as in the case of metals

REFERENCE

The Emission of Electricity from Hot Bodies. O. W. Richardson.

CHAPTER IV

Photo-electricity

1. Ultra-violet Light and Emission of Electrons.

Hertz in 1887 discovered that when ultra-violet light falls on a spark gap the sparks pass more easily. This effect was investigated by Hallwachs, Elster and Geitel, and others, and it was found that ultra-violet light causes a negatively charged conductor to lose its charge but has no effect on a positively charged conductor. Lenard made the important discovery that the kinetic energy of the negatively charged particles emitted when ultra-violet light falls on a negatively charged conductor is independent of the intensity of the light, but increases with the frequency of the light waves.

J. J. Thomson and Lenard in 1899 measured the ratio of the charge e to the mass m of the charged particles emitted, and found it equal to about 10^7 electromagnetic units per gram. This was about the same value of e/m as had been found for cathode rays or electrons, so that it was clear that ultra-violet light causes solid bodies to emit electrons.

The kinetic energy of the electrons emitted can be determined by finding the potential difference necessary to stop them.

An insulated electrode surrounded by a hollow metal conductor, both contained in a glass bulb which can be exhausted, is illuminated by the ultra-violet light. The electrode emits electrons and so becomes positively charged. The number of electrons which get across from the electrode to the surrounding conductor diminishes as the potential difference increases, and the potential difference soon attains a constant maximum value just sufficient to prevent any more electrons getting across. This maximum potential difference is a measure of the maximum kinetic energy of the electrons emitted. If V denotes this potential difference and e the charge on one electron, then Ve is equal to the maximum kinetic energy. The potential difference can be measured by connecting the electrode and conductor to an electroscope or quadrant electrometer. Experiments of this kind have been made by Ladenburg, A. Ll. Hughes, Richardson and Compton, and Millikan. Hughes, and Richardson and Compton found that Ve is of the same order of magnitude as $h\nu$, h being Planck's constant of the quantum theory and ν the frequency of the incident light. Millikan eliminated

various sources of error and obtained much more accurate results than previous workers, so that only his experiments need be considered in detail.

2. Millikan's Experiments. Critical Frequency.

In Millikan's experiments a small block of an alkali metal was illuminated by light in a very perfect vacuum. The surface of the block was scraped in the vacuum by a cutter worked by an electromagnet outside. In this way a clean metallic surface was obtained free from any film of oxide. The scraped surface was arranged in front of an insulated cylinder of oxidized copper gauze connected to a quadrant electrometer, and the charge received by this cylinder was observed when the surface was illuminated by monochromatic light of known frequency. The potential difference between the gauze cylinder and illuminated block was increased until it was enough to prevent any electrons from the block reaching the cylinder. The alkali metals emit electrons when exposed to light of much lower frequencies than are necessary to cause copper oxide to emit electrons. Millikan was therefore able to use only light which had no effect on the copper cylinder and so to avoid errors due to the cylinder emitting electrons. The contact potential difference between the alkali metal and copper oxide was determined by moving the metal block in the vacuum to a position opposite a plane copper electrode coated with oxide. This electrode could be moved in and out so as to vary the distance between it and the alkali metal block. The potential difference between this movable electrode and the block was adjusted until moving the electrode in and out made no difference to its potential when it was insulated. The potential difference is then equal to the contact potential difference and there is then no electric field in the space between the block and electrode. The potential difference required to stop the electrons from getting from the block to the gauze cylinder was taken to be the difference between the potential difference observed and the contact potential difference. In this way errors due to the contact potential difference were got rid of.

Millikan found that his results could be represented by the equation $Ve = h(\nu - \nu_0)$, where V is the corrected potential difference required to stop the electrons, h Planck's constant, ν the frequency of the light used, and ν_0 the smallest frequency which causes any electrons to be emitted. The frequency ν_0 is called the threshold or critical frequency and it is different for different substances. From his experiments Millikan obtained the value $h = 6{\cdot}56 \times 10^{-27}$, using the value of $e = 4{\cdot}774 \times 10^{-10}$ which he had previously determined. This value of h agrees well with that deduced from measurements of heat radiation and the frequencies of spectral lines, and is believed to be very near the true value.

It is found that the critical frequency for any substance depends

on the state of the surface used. Traces of gases and films of oxide or other impurities greatly affect it. The critical frequency is closely related to the contact potential difference and to the thermionic work function. Consider two parallel plates of different metals, and let them be connected together by a wire. If we disregard the small potential differences of thermoelectricity we may consider the two plates to be at the same potential. There will, nevertheless, be an electric field in the space between the plates equal to the contact potential difference divided by the distance between them. Let V_1 and V_2 be the potentials at the surface of the plates just outside the metals, and $e\phi_1$ and $e\phi_2$ the amounts of work required to remove an electron from inside the metals to a point just outside, then we have (see Chap. III, end of section 3).

$$\phi_1 - \phi_2 + V_2 - V_1 = 0.$$

3. Einstein's Theory.

According to the theory originally put forward by Einstein, when light falls on a metal electrons in the metal receive quanta $h\nu$ of energy from the light so that if they escape from the metal their kinetic energy is $h\nu - e\phi$, since $e\phi$ is the work required to remove an electron from the metal, ϕ being the thermionic work function.

Hence $\quad e\phi = h\nu_0,\;$ and $\; e(V_2 - V_1) = h(\nu_{02} - \nu_{01}) = e(\phi_2 - \phi_1).$

These relations appear to be confirmed by measurements of the contact potential differences, critical frequencies, and thermionic work functions, but cannot be regarded as definitely proved to be correct.

It appears that the electrons emitted with the maximum energy $h(\nu - \nu_0)$ must be the free electrons in the metal, since if electrons were emitted from the interior of atoms we should expect them to come out with energy less than $h\nu - e\phi$. Some of the electrons emitted may come from inside atoms, since most of them come out with less than the maximum energy. The light penetrates a short distance into the metal, so that some of the electrons may come out from points inside the metal and may lose energy by collisions before they get out.

Einstein's equation $Ve = h(\nu - \nu_0)$ holds good also for the emission of electrons due to X-rays.

The critical frequency ν_0 for most substances is in the ultra-violet region, but for the alkali metals it is in the visible spectrum or even in the infra-red.

The number of electrons emitted is found to be proportional to the intensity of the light.

4. Fermi-Dirac Theory.

Einstein's theory is not satisfactory because we should expect an electron in the metal with kinetic energy ϵ to escape with energy

$\epsilon + h\nu - \phi e$, so that the maximum energy of the escaping electrons should not have a definite value if the electrons have the Maxwell velocity distribution. This difficulty is removed by the Fermi-Dirac theory * of the electron gas in the metal.

According to this theory, at low temperatures, there are practically no free electrons with kinetic energies greater than a definite value

$$W = \frac{h^2}{2m}\left(\frac{3N}{8\pi}\right)^{2/3},$$

where N is the number of electrons per cm.^{3}. But there are a great many with all energies less than W. The maximum energy with which the photo-electrons are emitted should therefore have a definite value $W + h\nu - E$, where E is the least kinetic energy which the electrons must have in order to get out of the metal. If we put

$$W + h\nu - E = h\nu - \phi e, \text{ we get } E = W + \phi e.$$

Thus the least energy for escaping is not ϕe but $\phi e + W$. This result agrees with that found for thermionic emission, so we should expect ϕ for the photo-electric effect to have the same value as for the thermionic emission in agreement with the facts.

According to the Fermi-Dirac theory, there are a few electrons with energies greater than W, so that in the photo-electric effect there should be a corresponding small number emitted with energies greater than $W + h\nu - E$. It is found that this is the case, and the results agree with the Fermi-Dirac distribution.

5. Photo-electricity and the Classical Wave Theory.

It is apparently impossible to explain the phenomena of photo-electricity in a satisfactory manner by means of the classical wave theory of light. The difficulty is to explain how a very small fraction of the electrons get energy $h\nu$ and the rest get none. For example, when X-rays are passed through air a few atoms here and there emit electrons with energy $h\nu$ and the rest are apparently in no way affected.

According to the classical wave theory an electron which has a natural frequency of vibration equal to ν can absorb the energy of a train of waves, of frequency ν, from an area of the order of that of a square with sides one wave-length long. An ordinary X-ray tube gives out about 10^8 ergs of X-ray energy per second, and the ionization due to it can easily be detected at 3000 cm. The energy falling on one square wave length or about 10^{-15} sq. cm. at this distance from the tube is therefore about 10^{-15} ergs per second. The energy with which the electrons are emitted is about 10^{-7} ergs, so that it would take an electron 10^8 sec. or 1000 days to absorb it. But when the X-ray tube is started

* See p. 88

the ionization begins immediately. We might suppose that the energy of the emitted electrons is not derived from the rays but that the rays, so to speak, merely pull a trigger which releases the electron from the atom. However, if this were so we should expect the energy of the electrons emitted to be determined by the nature of the emitting atom, and not by the frequency of the rays as is the case. Moreover, it is difficult to believe that electrons having all possible natural frequencies are present in any substance. It seems certain that the energy of the electrons is derived from the rays, and the classical theory cannot explain how so much of it is concentrated into particular electrons in the time available. According to the quantum theory of radiation the radiation is supposed to consist of quanta, each of energy $h\nu$, which travel out from the source with the velocity of light. If an electron absorbs one of these quanta it gets energy $h\nu$. This quantum theory, therefore, explains the chief facts of photo-electricity, but of course it is hard to see how it can explain the facts of interference and diffraction which agree so well with the wave theory. This question is also discussed in the chapters on the quantum theory and on X-rays. Here we may say that there is no reason to suppose that a satisfactory explanation of both sets of facts will not eventually be discovered. It is quite unnecessary to adopt a mystical attitude and say that the human mind cannot understand such phenomena. Such an attitude is no satisfactory apology for an unintelligible theory.

5. Relation of Emission to Frequency. O. W. Richardson's Theory.

The number of electrons emitted by a metal per unit energy of the incident light depends upon the frequency of the light. It is zero for frequencies less than the critical frequency ν_0, and for frequencies greater than ν_0 it increases at first as ν increases, reaches a maximum value, and then decreases. The frequency which gives the maximum emission is about $\frac{3}{2}\nu_0$ in most cases. With some metals more than one maximum has been observed. An interesting theory of the variation of the emission with the frequency ν of the light has been given by O. W. Richardson. Consider a cylinder and piston maintained at a constant temperature T. The cylinder will be filled with black body radiation and this will cause the walls to emit electrons. Let the number of electrons per cubic centimetre in the cylinder be n and let the pressure they exert on the piston be p. We suppose n so small that the electric field due to the negative charges on the electrons may be neglected. If we suppose that the thermodynamical theory of evaporation can be applied to the escape of the electrons as in the theory of thermionics we have

$$N = \frac{2\pi mk^2}{h^3} T^2 \epsilon^{-w_0/kT},$$

where N is the number of electrons escaping from unit area in unit time, which is equal to the number absorbed from the electron gas by unit area in unit time. This equation is merely the equation for thermionic emission and the proof of it is given in the chapter on thermionics (Chap III, section 5).

Now let the energy density of the radiation between the frequencies ν and $\nu + d\nu$ be denoted by $E(\nu)d\nu$. The radiation falling on unit area in unit time is then $\frac{c}{4}E(\nu)d\nu$. For the proportion of E which belongs to rays whose directions lie within a small solid angle $d\omega$ is $d\omega/4\pi$. The radiation falling per unit time on a plane area A in directions inclined to the normal to A at angles between θ and $\theta + d\theta$ is therefore $Ec \cdot A \cos\theta \cdot 2\pi \sin\theta d\theta/4\pi$. Integrating from $\theta = 0$ to $\theta = \frac{\pi}{2}$, we get $\frac{1}{4}cEA$, the result stated. Let $F(\nu)$ be the number of electrons liberated from the surface by unit incident light energy of frequencies between ν and $\nu + d\nu$, so that the number of electrons liberated per unit area per unit time is

$$\frac{c}{4}\int_0^\infty F(\nu)E(\nu)\,d\nu = N.$$

For $E(\nu)$ we may use Wien's approximate formula (Chap. V, section 11) since only high frequencies are effective, so that

$$E(\nu) = \frac{8\pi}{c^3}h\nu^3\epsilon^{-h\nu/kT},$$

and we get $$\frac{c}{4}\int_0^\infty F(\nu)\frac{8\pi}{c^3}h\nu^3\epsilon^{-h\nu/kT}\,d\nu = \frac{2\pi mk^2}{h^3}T^2\epsilon^{-w_0/kT},$$

or $$\int_0^\infty F(\nu)\nu^3\epsilon^{-h\nu/kT}\,d\nu = \frac{c^2mk^2}{h^4}T^2\epsilon^{-w_0/kT}.$$

This equation is satisfied by

$$F(\nu) = \frac{mc^2}{\nu^2h^2}\left(1 - \frac{w_0}{h\nu}\right)$$

for $\nu > w_0/h$, and $F(\nu) = 0$ for $\nu < w_0/h$. Differentiating $F(\nu)$ with respect to ν and putting $\frac{dF(\nu)}{d\nu} = 0$, we find that $F(\nu)$ has a maximum value when $h\nu = \frac{3}{2}w_0$. This agrees with the experimental results that the maximum emission is obtained with a frequency equal to $\frac{3}{2}$ of the critical frequency, and that there is no emission when ν is less than the critical value w_0/h.

7. Thermionic Emission and Photo-electric Action.

In this theory the thermionic emission is not considered separately from the photo-electric emission, so that the theory really involves the assumption that the thermionic emission is caused by photo-electric action of radiation. Measurements of the photo-electric emission due to the radiation from hot tungsten by S. C. Roy (*Proc. Roy. Soc.*, Oct., 1926) show that it is quite possible that thermionic emission may be due entirely to photo-electric action.

The energy of the escaping electrons may be obtained if we assume that the electrons in the cylinder have the same energies as the molecules of a gas at the temperature T. The thermodynamical theory of evaporation would not be applicable to the electrons if this were not the case.

Let T_ν denote the kinetic energy of the electrons liberated by the light of frequency ν, so that

$$\frac{c}{4}\int_0^\infty T_\nu F(\nu) E(\nu) d\nu$$

is the total energy of the electrons escaping from unit area in unit time. This may be put equal to the energy of the N electrons falling on unit area in unit time, which is $2NkT$ according to the kinetic theory of gases. Using the value found for $F(\nu)$ we find that

$$T_\nu = h\nu - w_0$$

in agreement with Einstein's equation. This result is not in agreement with the experimental fact that most of the electrons emitted have energy less than $h\nu - w_0$, but the discrepancy may be only apparent. We suppose that the electrons receive energy $h\nu$ from the light initially, so that if they have energy less than $h\nu - w_0$ they must have lost some energy by collisions or otherwise.

Richardson's theory does not involve the assumption that the light consists of quanta having energy $h\nu$, but it is consistent with that view

When light of sufficiently high frequency is passed through gases electrons are liberated from the gas molecules. The critical frequencies for gases are greater than those for solid bodies. In the case of air the critical wave-length according to A. Ll. Hughes is about 1350 Å. The ionization of gases by X-rays is due to the emission of electrons by the atoms.

References

1. *Photo-electricity*. A. Ll. Hughes.
2. "Report on Photo-electricity." A. Ll. Hughes. (*Bulletin of National Research Council*, Washington, D.C.)

CHAPTER V

The Quantum Theory

1 Inadequacy of Newtonian Dynamics.

During the nineteenth century it was generally believed that material phenomena would prove capable of explanation on the old Newtonian system of dynamics. Matter was supposed to consist of minute particles which moved in accordance with Newton's laws of motion. However, certain phenomena, notably those of heat radiation, have emerged which seem to be inconsistent with the laws of classical dynamics, and the quantum theory was put forward by Max Planck as an explanation of such phenomena. This theory has now been applied successfully in several important branches of physics, for example, the theories of spectra, photo-electricity, and chemical equilibrium, and its fundamental character and value are universally recognized. The general validity of Newton's dynamics, therefore, can no longer be admitted. It appears that the laws of motion of bodies consisting of enormous numbers of atoms do not apply to atomic systems.

2. Microscopic and Macroscopic States. Statistical Mechanics.

It is necessary to distinguish between macroscopic and microscopic states of a substance. A macroscopic state is one determined by quantities such as pressure and temperature which can be measured by ordinary apparatus. For example, the macroscopic state of a gas is fixed by its pressure and volume. A microscopic state is one determined by the position and motion of all the parts of the substance, however small. Thus, a microscopic state of a gas requires for its specification the positions and velocities of all the molecules of which the gas is believed to be composed.

Quantities such as pressure, internal energy per unit mass, and temperature, which are used to specify the condition of a substance, are average values over very large numbers of atoms. If, for example, we know the temperature and the entropy of a given quantity of any homogeneous substance, then we know that it is in a certain macroscopic state capable of being reproduced and indistinguishable from

other states of the same substance having equal temperatures and specific entropies. But for any such macroscopic state there is an enormous number of different microscopic states. Thus, a gas at a given temperature and pressure is in a definite macroscopic state, but its molecules are moving about so that its microscopic state is continually changing, and in a short time it passes through an enormous number of different microscopic states. The heat radiation inside an empty enclosure the walls of which are maintained at a constant temperature is in a definite macroscopic state. The energy density for any given range of frequencies has a definite value depending only on the temperature of the walls. But to specify the microscopic state of the radiation in the enclosure at any instant it would be necessary to give the strength and direction of the electric and magnetic fields at every point in the enclosure. The radiation in the enclosure is continually moving with the velocity of light, so that its microscopic state is continually changing and it passes through an enormous number of microscopic states in a short time.

It is clear that to any macroscopic state of a substance there corresponds an enormous, perhaps an infinite, number of microscopic states. During the rapid change from one microscopic state to another which continually goes on in material substances which are in a state of equilibrium in a fixed macroscopic state, the macroscopic quantities determining the macroscopic state do not vary perceptibly. The value of such a macroscopic quantity, which can be measured, is an average over a large number of molecules for a short but finite time, or over a large volume, and it remains constant within the limits of error of observation in ordinary cases.

If a solid sphere is immersed in a gas, the uniform pressure of the gas over the surface of the sphere gives no resultant force on the sphere when we consider the action of the gas on the sphere from the macroscopic or large-scale point of view. Microscopically, however, the sphere is not subjected to a uniform pressure but to a series of molecular impacts, and, indeed, if the sphere is very small it does not remain at rest but moves about in an irregular manner, owing to the irregular distribution of the molecular impacts over its surface. This motion, the well-known Brownian movements of small particles immersed in a liquid or gas, is a microscopic motion. Macroscopically, the sphere is to be regarded as at rest in its mean position.

The question arises how it is that the macroscopic state of a substance can remain sensibly constant while its microscopic state is continually changing. The reason must be that of all the possible microscopic states the vast majority correspond to values of the macroscopic quantities differing inappreciably from the constant values. If any microscopic state have an appreciably different value of the macroscopic quantities, the chance of its lasting long enough for the

change in a macroscopic quantity to be observable is therefore negligible.

When a substance is not in a state of macroscopic equilibrium its state changes until equilibrium is reached. The relation of such changes to the microscopic states will now be considered. Since the number of microscopic states corresponding to any state differing appreciably from the state of macroscopic equilibrium must be very small compared with the whole number of possible microscopic states, the chance of the substance remaining an appreciable time in such a state must also be very small, and the substance changes to the equilibrium macroscopic state because this state corresponds to nearly all possible microscopic states. The change from a state which is not one of equilibrium to an equilibrium state thus involves an increase in the number of possible microscopic states. It is clear, therefore, that an equilibrium state is one for which the number of microscopic states is a maximum. Such considerations, of course, only apply to substances or systems the microscopic parts of which are continually changing from one microscopic state to another, so that in the course of time the system may be supposed to pass through all possible microscopic states.

It appears then that when a substance consists of a very great number of individuals, the behaviour of the substance as a whole must be investigated by methods of statistics.

3. Entropy and Probability.

Let W denote the number of possible microscopic states through which the substance may pass in the course of time while in a given macroscopic state. Then the condition of equilibrium is $W=$ maximum or $\delta W=0$. If the microscopic parts of the system are at rest so that its microscopic state does not change with time then $W=1$. According to the second law of thermodynamics the entropy Φ of any isolated system is increased by any spontaneous change which takes place in the system, and in a state of equilibrium the entropy is a maximum. Thus it appears that both W and Φ tend to increase to maximum values. W may be called the thermodynamical probability of the state of the system, and we may then say that the system changes to a more probable state unless it is in a state of maximum probability. The second law of thermodynamics evidently expresses the same thing in terms of Φ instead of W, so that we should expect Φ to depend on W. This idea was first put forward by Boltzmann. Let us then suppose that the entropy Φ is some function of W, or let

$$\Phi=f(W).$$

Now consider two entirely separate systems having entropies Φ_1 and Φ_2 and thermodynamic probabilities W_1 and W_2. Then

$$\Phi_1=f(W_1),\quad \Phi_2=f(W_2).$$

The total entropy of the two systems is $\Phi_1 + \Phi_2$, and since for each microscopic state of the first system there are W_2 states of the second, the thermodynamic probability of the two systems considered together as one must be the product $W_1 W_2$. Thus we have

$$\Phi_1 + \Phi_2 = f(W_1 W_2)$$

so that

$$f(W_1) + f(W_2) = f(W_1 W_2).$$

Differentiating this equation with respect to W_1, and also with respect to W_2, we obtain

$$W_1 f'(W_1) = W_2 f'(W_2) = W_1 W_2 f'(W_1 W_2).$$

We may suppose W_1 to change while W_2 remains constant, since the two systems were supposed entirely separate, so that we must have

$$W_1 f'(W_1) = W_2 f'(W_2) = \text{constant}.$$

Hence $f(W) = k \log W + C$ where k is a universal constant, the same for all systems, and C is another constant, not necessarily the same for all systems. Hence $\Phi = k \log W + C$ for any system. In classical thermodynamics the entropy is taken to be the difference between the entropy in the actual state and that in an arbitrarily chosen standard state. If W_0 is the thermodynamical probability in the standard state then $\Phi = k \log W - k \log W_0 = k \log (W/W_0)$.

The number of possible microscopic states W of a system corresponding to a given macroscopic state might be expected to be infinite. Thus if the velocity of an atom can vary continuously it has an infinite number of possible values, so that a gas at a given pressure and temperature should have an infinite number of possible microscopic states. But if W is infinite then $\Phi = k \log W + C$ must also be infinite. We might suppose that though W and W_0 were both infinite their ratio W/W_0 could have a definite finite value, but this method of avoiding the difficulty has not proved satisfactory. Planck supposes that W is not infinite but has a definite value for any system, and also that the constant C is equal to zero so that $\Phi = k \log W$. This leads to an absolute value of the entropy and to the quantum theory. It follows that a system in a given macroscopic state can only exist in a finite number of microscopic states. Thus, for example, the velocities of the molecules of a gas cannot vary continuously, but must be restricted in some way to a finite number of possible values, or at least changes in the velocities less than certain finite amounts must be supposed not to constitute a change in the microscopic state.

4. State Space of a System.

Consider a system of any kind, for example, an atom of any element. Let the microscopic state of the system be determined by co-ordinates $q_1, q_2, \ldots q_n$ and the corresponding momenta $p_1, p_2, \ldots p_n$, that is, let p_m be the momentum associated with the co-ordinate q_m. The $2n$ quantities $p_1, p_2, \ldots p_n$ and $q_1, q_2, \ldots q_n$ may be regarded as the co-ordinates of a point in a space of $2n$ dimensions. This representative point will move along a path which in the case of a periodic motion of the system will be a closed curve. So long as the energy of the system remains unchanged the point will continue to move round and round this closed curve, and if we regard the state of the system as determined by its energy then the motion of the representative point round the curve does not involve a change in the state of the system. The space of $2n$ dimensions is called the state space of the system.

Suppose now that we have a large number $\mathcal{N}$ of such systems all alike so that the state of each one can be represented in the space of $2n$ dimensions by a representative point. The $\mathcal{N}$ points will all describe curves in the space. The microscopic state of the collection may be regarded as determined by the distribution of energy among the $\mathcal{N}$ systems. In the state space we can imagine surfaces or regions drawn in which the energy is constant. Then all the systems having the same energy will have representative points moving on the same surface of constant energy. According to classical dynamics we should expect that the energy of the system could vary continuously, so that there would be an infinite number of regions or surfaces of constant energy on which the representative points could move. The number of possible microscopic states of the collection of $\mathcal{N}$ systems would therefore be infinite. In the quantum theory it is supposed that the number of possible microscopic states is finite, so that only certain definite values of the energy of a system are possible. The surfaces of constant energy corresponding to these possible values will divide the state space up into finite regions, and all the representative points will be on these surfaces.

When the energy of one of the systems changes from one possible value to another, its representative point is supposed to jump from one of the surfaces to another without occupying the intermediate positions effectively. Such jumps are supposed to be due to actions between the systems or between the systems and the radiation in the collection of $\mathcal{N}$ systems. It is important to remember that any collection of atoms always contains a certain amount of energy in the form of radiation. This radiant energy is generally small compared with the energy of the atoms and so can be neglected in many cases, but it provides a reservoir of energy into which or from which the atoms can give out or absorb energy.

5. Planck's Theory of Entropy and Free Energy.

Let $\epsilon_1, \epsilon_2, \epsilon_3 \ldots$ denote the possible values of the energy of a system, so that the total energy of the $\mathcal{N}$ similar systems or parts of the collection is given by

$$E = N_1\epsilon_1 + N_2\epsilon_2 + \ldots = \Sigma N\epsilon,$$

where $N_1, N_2 \ldots$ denote the numbers of the parts having energies ϵ_1, $\epsilon_2 \ldots$ respectively.

If two systems having energies ϵ_1 and ϵ_2 act on each other so that their energies change to ϵ_1' and ϵ_2', then if $\epsilon_1 + \epsilon_2$ is not equal to $\epsilon_1' + \epsilon_2'$ the difference may be supposed supplied from the radiation present. Also if the energy of a system changes spontaneously from ϵ_1 to ϵ_1' the difference $\epsilon_1 - \epsilon_1'$ may be supposed emitted in the form of radiation. The total radiant energy is small compared with E, and its average value remains constant so long as E is not changed, but it must fluctuate about its average value owing to the emission and absorption as the systems change from one possible value of ϵ to another. The number W of possible microscopic states corresponding to a given distribution of the energy is given by

$$W = \frac{\mathcal{N}!}{N_1!N_2!N_3! \ldots}.$$

For if all the $\mathcal{N}$ parts got different amounts of energy the number of possible arrangements would be equal to the number of permutations of $\mathcal{N}$ different objects, each permutation containing all the objects, or $\mathcal{N}!$. When N_1 of the objects are identical, let the number of different permutations be x. By changing the N_1 identical objects to N_1 different objects, $N_1!$ different permutations could be made out of each of the x permutations, so that $$xN_1! = \mathcal{N}!.$$

In the same way it is easy to see that when the objects are divided into groups containing N_1, N_2, &c., identical objects the number of permutations is equal to the expression given above for W.

An approximate value of $N!$, sufficiently exact for the purpose of calculating $\log W$ when $\mathcal{N}$, N_1, N_2, &c., are large, may be easily obtained. We have

$$\log N! = \log 1 + \log 2 + \log 3 + \log 4 + \ldots + \log N,$$

and $$\log N^N = \log N + \log N + \ldots + \log N,$$

so that $$\log (N!/N^N) = \log \frac{1}{N} + \log \frac{2}{N} + \ldots + \log \frac{N}{N}.$$

Put $\frac{1}{N} = dx$, so that

$$\frac{1}{N}\log(N!/N^N) = \log dx\cdot dx + \log 2dx\cdot dx + \ldots + \log N dx\cdot dx$$
$$= \int_0^1 \log x\cdot dx = -1.$$

Hence $\log N! = N\log N - N = \log(N/e)^N$, so that $N! = \left(\frac{N}{e}\right)^N$. Using this value for $N!$, we get

$$W = \frac{(\mathcal{N}/e)^{\mathcal{N}}}{(N_1/e)^{N_1}(N_2/e)^{N_2}\ldots} = \frac{\mathcal{N}^{\mathcal{N}}}{N_1^{N_1}N_2^{N_2}\cdot N_3^{N_3}\ldots},$$

since $N_1 + N_2 + \ldots = \mathcal{N}$.

Hence $$\log W = \mathcal{N}\log\mathcal{N} - \Sigma N\log N.$$

Now let $w_1 = N_1/\mathcal{N}$, $w_2 = N_2/\mathcal{N}$, &c., so that $\Sigma w = 1$, and $\log W = \mathcal{N}\log\mathcal{N} - \Sigma\mathcal{N}w\log\mathcal{N}w$,

or $$\log W = -\mathcal{N}\Sigma w\log w.$$

The entropy Φ of the system is then given by

$$\Phi = k\log W = -k\mathcal{N}\Sigma w\log w.$$

To find the distribution of the energy in the system in the state of equilibrium, we make Φ a maximum subject to the conditions $\Sigma w = 1$ and $E = \mathcal{N}\Sigma w\epsilon =$ constant.

Thus $$\delta\Phi = 0 = \Sigma\delta w\log w + \Sigma\delta w = \Sigma\delta w\log w,$$
$$\delta E = 0 = \Sigma\epsilon\delta w,\ \text{and}\ \Sigma\delta w = 0.$$

Hence $$\Sigma(\log w + \beta\epsilon + \gamma)\delta w = 0,$$

where β and γ are undetermined multipliers. If β and γ are properly chosen, this equation will be true for any values of the δw's, so that we have

$$\log w_n = -\beta\epsilon_n - \gamma,$$

or $$w_n = \alpha e^{-\beta\epsilon_n}. \quad \ldots\ldots\ldots \quad (1)$$

Here α and β are constants having the same values for all parts of the system. The entropy in the equilibrium state is therefore given by

$$\Phi = k\mathcal{N}\Sigma w(\beta\epsilon - \log\alpha),$$

or $$\Phi = k\mathcal{N}\beta\Sigma w\epsilon - k\mathcal{N}\log\alpha,$$

so that $$\Phi = k\beta E + k\mathcal{N}\log\Sigma e^{-\beta\epsilon_n}, \quad \ldots\ldots \quad (2)$$

since $$\Sigma w_n = \alpha\Sigma e^{-\beta\epsilon_n} = 1,\ \text{and}\ E = \mathcal{N}\Sigma w_n\epsilon_n.$$

The thermodynamical definition of Φ gives

$$\delta\Phi = \frac{\delta E + p\,\delta V}{\theta},$$

where p, V, and θ denote the pressure, volume, and absolute temperature of the system. But, regarding Φ as a function of E and V, we have

$$\delta\Phi = \left(\frac{\partial\Phi}{\partial E}\right)_V \delta E + \left(\frac{\partial\Phi}{\partial V}\right)_E \delta V,$$

so that

$$\left(\frac{\partial\Phi}{\partial E}\right)_V = \frac{1}{\theta} \text{ and } \left(\frac{\partial\Phi}{\partial V}\right)_E = \frac{p}{\theta}.$$

Differentiating the equation $\Phi = k\beta E + k\mathcal{N} \log \Sigma e^{-\beta\epsilon_n}$ with respect to β, we get

$$\frac{\partial\Phi}{\partial\beta} = kE + k\beta\frac{\partial E}{\partial\beta} - \frac{k\mathcal{N}\Sigma\epsilon_n e^{-\beta\epsilon_n}}{\Sigma e^{-\beta\epsilon_n}}.$$

But

$$\frac{\mathcal{N}\Sigma\epsilon_n e^{-\beta\epsilon_n}}{\Sigma e^{-\beta\epsilon_n}} = \mathcal{N}a\Sigma\epsilon_n e^{-\beta\epsilon_n} = \mathcal{N}\Sigma w_n\epsilon_n = E,$$

so that

$$\frac{\partial\Phi}{\partial\beta} = k\beta\frac{\partial E}{\partial\beta}.$$

Hence

$$\frac{\partial\Phi}{\partial E} = \frac{\partial\Phi}{\partial\beta}\,\frac{\partial\beta}{\partial E} = k\beta = \frac{1}{\theta}.$$

This enables us to introduce the temperatures into the expression (2) for the entropy, and obtain

$$\Phi = \frac{E}{\theta} + k\mathcal{N} \log \Sigma e^{-\epsilon_n/k\theta}. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The free energy F of the system is defined to be $E - \theta\Phi$, so that

$$F = -\,k\mathcal{N}\theta \log \Sigma e^{-\epsilon_n/k\theta}.$$

This gives

$$\theta\frac{\partial F}{\partial\theta} = -\,k\mathcal{N}\theta \log \Sigma e^{-\epsilon_n/k\theta} - \frac{k\mathcal{N}\theta^2}{k\theta^2}\,\frac{\Sigma\epsilon_n e^{-\epsilon_n/k\theta}}{\Sigma e^{-\epsilon_n/k\theta}},$$

or

$$E = F - \theta\frac{\partial F}{\partial\theta}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

This is the well-known Gibbs-Helmholtz equation, from which the thermodynamical theory of the system may be deduced. Thus we see that the quantum theory is consistent with classical thermodynamics.

Since $E = F + \theta\Phi$, equation (4) gives $\Phi = -\dfrac{\partial F}{\partial\theta}$.

6. Planck's Constant.

In order to use (3) to calculate the entropy of a system it is of course necessary to determine the possible values ϵ_n of the energies of the $\mathcal{N}$ independent parts This may require some knowledge of the physical properties of the parts and involves further assumptions characteristic of the quantum theory. The surfaces of constant energy for the possible values of the energy divide the state space into certain finite regions. In the case of periodic systems determined by only one co-ordinate q and the corresponding momentum p, the state space is a plane and the closed curves of possible constant energies are supposed by Planck to divide the plane into equal areas h. The area of the closed curve of energy ϵ_n is then equal to nh, so that

$$\int dpdq = \int pdq = nh,$$

where $n = 0, 1, 2, 3, \ldots$.

The quantity h is called *Planck's constant.* It appears to have the same value in all such cases, and its value has been determined by comparing results deduced from the quantum theory with experimental values. The most probable value of h is $6 \cdot 55 \times 10^{-27}$ erg-seconds. The dimensions of h are (energy) $\times$ (time), or action, for

$$(m\dot{x})\,dx = (m\dot{x}^2)\,dt.$$

7. Monatomic Gas.

As an example, consider the case of $\mathcal{N}$ atoms of a monatomic gas, each of mass m, and each contained in a separate cubical box of volume V. Let the co-ordinates of an atom measured along the edges of its box be x, y, z, and the corresponding momenta $m\dot{x}$, $m\dot{y}$, and $m\dot{z}$ The state space will be of six dimensions, and the representative points of all the atoms having kinetic energies equal to s will lie on the surface of a sphere of radius r given by

$$r^2 = (m\dot{x})^2 + (m\dot{y})^2 + (m\dot{z})^2,$$
$$s = \tfrac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2),$$

so that

$$r = \sqrt{2sm},$$

and they all also lie inside the volume V. The six-dimensional volume enclosed by the surface of constant kinetic energy s is therefore equal to

$$\tfrac{4}{3}\pi(2sm)^{3/2}V.$$

We shall suppose that the surfaces of constant energy corresponding to possible values of s divide the state space into equal regions h^3, since in this case each system has three co-ordinates. Hence

$$\tfrac{4}{3}\pi(2s_n m)^{3/2}V = nh^3,$$

where $n = 0, 1, 2, 3, \ldots$.

The total energy of an atom ϵ_n may be taken equal to $s_n + \epsilon_0$, where ϵ_0 denotes its internal energy. The free energy of the collection of $\mathcal{N}$ atoms in the equilibrium state of maximum entropy can now be calculated by means of the equation

$$F = -k\mathcal{N}\theta \log \Sigma e^{-\epsilon_n/k\theta}.$$

The values of s_n are very small unless n is very large, so that unless the total energy E of the $\mathfrak{N}$ atoms is very small we may replace $\Sigma e^{-\epsilon_n/k\theta} = e^{-\epsilon_0/k\theta}\Sigma e^{-s_n/k\theta}$ by an integral

We have
$$s_n = \frac{h^2}{2m}\left(\frac{3n}{4\pi V}\right)^{2/3} = \alpha k\theta\, n^{2/3},$$
where α is a constant. Then
$$\Sigma e^{-s_n/k\theta} = \Sigma e^{-\alpha n^{2/3}} = \int_0^\infty e^{-\alpha n^{2/3}} dn.$$

Put $x = n^{1/3}$ so that $dx = \frac{1}{3}n^{-2/3}dn$, and
$$\int_0^\infty e^{-\alpha n^{2/3}}dn = 3\int_0^\infty x^2 e^{-\alpha x^2}dx = \tfrac{3}{4}\sqrt{\frac{\pi}{\alpha^3}}.$$

Hence
$$\Sigma e^{-\epsilon_n/k\theta} = \frac{(2\pi mk\theta)^{3/2} V e^{-\epsilon_0/k\theta}}{h^3}.$$

The free energy of the $\mathfrak{N}$ atoms in separate boxes is therefore given by
$$F = -k\mathfrak{N}\theta \log\{(2\pi mk\theta)^{3/2} V/h^3\} + \mathfrak{N}\epsilon_0,$$

and the entropy $\Phi = -\partial F/\partial\theta$ by
$$\Phi = k\mathfrak{N}\log\{(2\pi mk\theta)^{3/2} V/h^3\} + \tfrac{3}{2}k\mathfrak{N}. \quad . \quad . \quad . \qquad (5)$$

The energy E of the $\mathfrak{N}$ atoms is then given by
$$E = F - \theta\frac{\partial F}{\partial\theta} = \mathfrak{N}(\tfrac{3}{2}k\theta + \epsilon_0),$$

so that it appears that $\frac{3}{2}k\theta$ is the average kinetic energy of one atom. If we now suppose the $\mathfrak{N}$ atoms all contained in one box of volume V, then if the $\mathfrak{N}$ atoms are all different the number of ways in which the energy elements can be arranged among the atoms will be the same as before, so that the entropy will still be given by equation (5).

If the $\mathfrak{N}$ atoms are all different, the number of ways in which they can be arranged among $\mathfrak{N}$ quantities of energy is $\mathfrak{N}!$, so that when they are all alike and it makes no difference which particular atom has a given amount of energy the number of possible arrangements is diminished $\mathfrak{N}!$ times, and $k \log \mathfrak{N}!$, or $k\mathfrak{N} \log\left(\frac{\mathfrak{N}}{e}\right)$, must be subtracted from the entropy given by (5). Hence we get for the entropy of a monotomic gas
$$\Phi = k\mathfrak{N}\log\left\{\frac{Ve^{5/2}}{\mathfrak{N}h^3}(2\pi mk\theta)^{3/2}\right\}.$$

Differentiating this with respect to V, keeping E and therefore θ constant, we get
$$\left(\frac{\partial\Phi}{\partial V}\right)_E = \frac{p}{\theta} = \frac{k\mathfrak{N}}{V},$$
so that $pV = k\mathfrak{N}\theta$.

This equation expresses the well-known relation between the pressure, volume, and temperature of a gas, and shows that k is equal to the gas constant for one molecule The energy of the gas is the same as that of the $\mathfrak{N}$ atoms in separate boxes, so that
$$E = \tfrac{3}{2}k\mathfrak{N}\theta + \mathfrak{N}\epsilon_0,$$

and the free energy is given by

$$F = E - \theta\Phi = -k\mathcal{N}\theta \log\left\{\frac{Ve}{\mathcal{N}h^3}(2\pi mk\theta)^{3/2}\right\} + \mathcal{N}\varepsilon_0.$$

The distribution of the energy among the atoms of gas is given by the equation

$$w_n = \alpha e^{-\varepsilon_n/k\theta},$$

which when $\varepsilon_{n+1} - \varepsilon_n$ is very small expresses Maxwell's law. Thus we have obtained the principal results of the kinetic theory of gases by means of the quantum theory.

8. Vapour Pressure.

The latent heat of evaporation, L, of any solid or liquid at a constant temperature θ is given by the equation

$$\frac{L}{\theta} = \Phi_1 - \Phi_2,$$

where Φ_1 denotes the entropy of the vapour and Φ_2 that of the solid or liquid. At low temperatures Φ_2 is small and the vapour may be regarded as a perfect gas, so that for substances giving monatomic vapours we have

$$\frac{L}{\theta} = k\mathcal{N} \log\left\{\frac{Ve^{5/2}}{\mathcal{N}h^3}(2\pi mk\theta)^{3/2}\right\},$$

where L is the heat of evaporation of $\mathcal{N}$ atoms. Putting $V = k\mathcal{N}\theta/p$ and solving for $\log p$ we get

$$\log p = -\frac{L}{k\mathcal{N}\theta} + \log\left\{\left(\frac{2\pi m}{h^2}\right)^{3/2}(ke\theta)^{5/2}\right\}.$$

This equation has been found to agree very well with the observed vapour pressures of mercury, argon, helium, and other monatomic substances at low temperatures when the values of h and k derived from experiments on heat radiation are substituted in it. It may be written

$$\log p = -\frac{L}{k\mathcal{N}\theta} + \tfrac{5}{2}\log\theta + \tfrac{5}{2} + i, \quad \ldots\ldots \quad (6)$$

where

$$i = \log\left\{\left(\frac{2\pi}{h^2}\right)^{3/2}k^{5/2}\right\} + \tfrac{3}{2}\log m$$

is Nernst's "chemical constant" for a monatomic substance. The pressure of electron gas in equilibrium with a hot metal may also be calculated by (6) if the heat energy absorbed when $\mathcal{N}$ electrons escape from the metal is substituted for L, and for m the mass of one electron. This important application of the quantum theory is discussed in the chapter on thermionics (Chap. III, section 5).

9. Simple Oscillators.

We will now consider the case of a system of a large number $\mathcal{N}$ of oscillators, each consisting of a particle of mass m, which can only move along a straight line under the action of a force proportional to its distance from a fixed point in the line and directed toward the fixed point. If x denotes the distance of the particle from the fixed point, then $m\ddot{x} = -\mu x$, where μ is a constant. A solution of this equation is

$x = A \sin 2\pi\nu t$, where A is the amplitude of the oscillations of the particle and ν the frequency or number of vibrations in unit time. Substituting this value of x in the equation $m\ddot{x} = -\mu x$ we find

$$\nu = \frac{1}{2\pi}\sqrt{\frac{\mu}{m}}.$$

The energy ϵ of the vibrating particle is given by

$$\epsilon = \tfrac{1}{2}m\dot{x}^2 + \tfrac{1}{2}\mu x^2 + \epsilon_0,$$

where ϵ_0 stands for any energy it may have when at rest in its equilibrium position. Putting $m\dot{x} = y$, we get

$$1 = \frac{y^2}{2m(\epsilon - \epsilon_0)} + \frac{\mu x^2}{2(\epsilon - \epsilon_0)}.$$

This relation between x and y is represented by an ellipse in the xy plane. The representative point of the oscillator moves round the ellipse, making ν revolutions in unit time. The area of the ellipse is

$$2\pi(\epsilon - \epsilon_0)\sqrt{\frac{m}{\mu}},$$

which is equal to $(\epsilon - \epsilon_0)/\nu$.

According to the quantum theory the possible paths of the particle in the xy plane will divide the plane into parts of equal area h. The possible paths will be a series of ellipses having areas 0, h, $2h$, $3h$, &c., so that the possible values of the energy are given by $\epsilon - \epsilon_0 = nh\nu$, where $n = 0, 1, 2, 3$, &c. The free energy of the system of $\mathcal{N}$ oscillators when the energy is distributed so that the entropy has the maximum value is given (section 5) by $F = -k\mathcal{N}\theta \log \Sigma e^{-\epsilon_n/k\theta}$.

We have $\quad \Sigma e^{-\epsilon_n/k\theta} = e^{-\epsilon_0/k\theta}\{1 + e^{-h\nu/k\theta} + e^{-2h\nu/k\theta} + \dots\}$

$$= \frac{e^{-\epsilon_0/k\theta}}{1 - e^{-h\nu/k\theta}},$$

so that $$F = k\mathcal{N}\theta \log(1 - e^{-h\nu/k\theta}) + \mathcal{N}\epsilon_0.$$

Hence the total energy E of the $\mathcal{N}$ oscillators in the state of maximum entropy is equal to

$$E = F - \theta\frac{\partial F}{\partial \theta} = \frac{\mathcal{N}h\nu}{e^{h\nu/k\theta} - 1} + \mathcal{N}\epsilon_0.$$

The average energy of vibration, $\epsilon - \epsilon_0$, of one oscillator is therefore $h\nu/(e^{h\nu/k\theta} - 1)$, and the fraction w_n of the $\mathcal{N}$ oscillators which have energy $\epsilon_n - \epsilon_0 = nh\nu$ is equal to $e^{-\epsilon_n/k\theta}/\Sigma e^{-\epsilon_n/k\theta}$, or $e^{-nh\nu/k\theta}(1 - e^{-h\nu/k\theta})$. The fraction which has no energy is $1 - e^{-h\nu/k\theta}$. At high temperatures when $h\nu/k\theta$ is small the average vibrational energy per oscillator is

$k\theta$, since $e^{h\nu/k\theta} - 1 = h\nu/k\theta$ when $h\nu/k\theta$ is small, and the fraction having zero vibrational energy is $h\nu/k\theta$, which is small. At low temperatures when $h\nu/k\theta$ is large the average energy is much less than $h\nu$, and nearly all the oscillators have zero vibrational energy. The heat capacity of the system of oscillators is given by

$$\frac{dE}{d\theta} = \frac{\mathcal{N}(h\nu)^2 e^{h\nu/k\theta}}{k\theta^2(e^{h\nu/k\theta} - 1)^2}.$$

This makes the heat capacity very small at low temperatures and equal to $\mathcal{N}k$ at high temperatures.

It appears that the average energy per oscillator in the equilibrium state of maximum entropy depends only on the temperature and the frequency of the oscillators. It is easy to see that this result will apply to any collection of oscillators, even if they are not all similar, provided the energy of each is determined by one co-ordinate and the corresponding momentum. Thus if we consider a solid body and suppose that it possesses $\mathcal{N}_1$ modes of vibration of frequency ν_1, $\mathcal{N}_2$ of frequency ν_2, &c., the energy of its vibrations will be equal to

$$\frac{\mathcal{N}_1 h\nu_1}{e^{h\nu_1/k\theta} - 1} + \frac{\mathcal{N}_2 h\nu_2}{e^{h\nu_2/k\theta} - 1} + \ldots = \Sigma \frac{\mathcal{N}_n h\nu_n}{e^{h\nu_n/k\theta} - 1}.$$

10. Quantum Theory of Specific Heat.

In this way Einstein and Debye have worked out a quantum theory of the specific heats of solid bodies. We may regard the heat energy of the solid as the energy of elastic waves travelling through it like sound waves through air. The possible frequencies of vibration are the frequencies of stationary waves possible in the solid.

Consider a cube of the solid with sides of length a, and let AB (fig. 1) be a plane wave travelling in the direction CN. Draw a plane $A'B'$ at a distance from AB of one-half wave-length. Take the origin O at one corner of the cube and axes x, y, z along its edges. Let the cosines of the angles between the direction of propagation of the plane wave and the axes be l, m, n. Let EF be a line parallel to the y axis cutting the planes AB and $A'B'$ at C and C'. Then

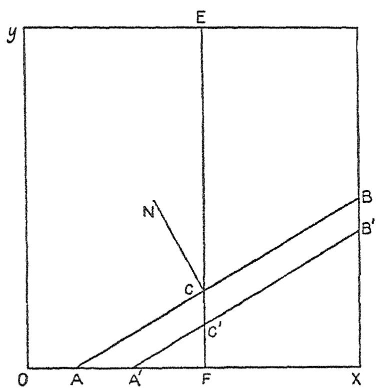


Fig 1

$$CC' = \frac{\lambda}{2m}.$$

If the wave AB is to be one of the systems of standing waves of length λ we must have $n_2 CC' = a$, where n_2 is a positive integer. Hence

$$\frac{n_2\lambda}{2m} = a.*$$

In the same way, considering lines like EF parallel to the x and z axes we get

$$\frac{n_1\lambda}{2l} = a, \text{ and } \frac{n_3\lambda}{2n} = a.$$

Hence $$l^2 + m^2 + n^2 = 1 = \left(\frac{\lambda}{2a}\right)^2 (n_1^2 + n_2^2 + n_3^2).$$

If v is the velocity of propagation of the waves and ν the frequency, then $\nu\lambda = v$, so that

$$n_1^2 + n_2^2 + n_3^2 = \left(\frac{2a}{v}\nu\right)^2.$$

Now let n_1, n_2, n_3 be the rectangular co-ordinates of points, so that there will be one point in each unit volume. The number of possible frequencies between 0 and ν is thus equal to one-eighth of the volume of a sphere of radius $\frac{2a\nu}{v}$, or $\frac{1}{8}\cdot\frac{4}{3}\pi\left(\frac{2a\nu}{v}\right)^3$, and the number of possible frequencies between 0 and ν per unit volume is equal to $\frac{4}{3}\pi\left(\frac{\nu}{v}\right)^3$.

In a solid there will be longitudinal waves and also transverse waves, and the latter can be regarded as forming two sets polarized in perpendicular planes. The total number of frequencies between ν and $\nu + d\nu$ in an elastic solid is therefore

$$4\pi\nu^2\left(\frac{2}{v_1^3} + \frac{1}{v_2^3}\right)d\nu,$$

where v_1 is the velocity of transverse, and v_2 that of longitudinal waves.

The energy of the vibrations in unit volume between the frequencies 0 and ν is therefore

$$4\pi h\left(\frac{2}{v_1^3} + \frac{1}{v_2^3}\right)\int_0^{\nu} \frac{\nu^3 d\nu}{e^{h\nu/k\theta} - 1}.$$

If the solid contains $\mathcal{N}$ atoms in unit volume their positions could be determined by $3\mathcal{N}$ co-ordinates, so that we should not expect more than $3\mathcal{N}$ possible modes of vibration per unit volume in the solid. Debye therefore supposes that the maximum possible frequency ν_m is given by

$$3\mathcal{N} = 4\pi\left(\frac{2}{v_1^3} + \frac{1}{v_2^3}\right)\int_0^{\nu_m} \nu^2 d\nu = \frac{4}{3}\pi\left(\frac{2}{v_1^3} + \frac{1}{v_2^3}\right)\nu_m^3.$$

* Cf. J. K. Roberts, *Heat and Thermodynamics*, Chap. XX, section 2.

The total energy E of the vibrations in unit volume is then given by

$$E = 4\pi h\left(\frac{2}{v_1^3} + \frac{1}{v_2^3}\right)\int_0^{\nu_m}\frac{\nu^3 d\nu}{e^{h\nu/k\theta} - 1}.$$

If we put $h\nu/k\theta = x$, and $\psi = h\nu_m/k$, this equation for E becomes

$$E = 9\mathcal{N}k\frac{\theta^4}{\psi^3}\int_0^{\psi/\theta}\frac{x^3 dx}{e^x - 1}.$$

The heat capacity of $\mathcal{N}$ atoms is given by

$$\frac{\partial E}{\partial \theta} = 9\mathcal{N}k\left\{4\left(\frac{\theta}{\psi}\right)^3\int_0^{\psi/\theta}\frac{x^3 dx}{e^x - 1} - \frac{\psi}{\theta}\,\frac{1}{e^{\psi/\theta} - 1}\right\}. \quad . \quad . \quad (7)$$

When θ is very large this reduces to $\dfrac{\partial E}{\partial \theta} = 3\mathcal{N}k$, which agrees with Dulong and Petit's law. At low temperatures ψ/θ becomes large, so that

$$\frac{\partial E}{\partial \theta} = 36\mathcal{N}k\left(\frac{\theta}{\psi}\right)^3\int_0^{\infty}\frac{x^3 dx^*}{e^x - 1} = \tfrac{12}{5}\pi^4\mathcal{N}k\left(\frac{\theta}{\psi}\right)^3. \quad . \quad . \quad (8)$$

Thus at low temperatures the atomic heat is proportional to the cube of the absolute temperature. Equation (7) shows that the atomic heat capacity is the same function of θ/ψ for all solid substances.

ψ is called the *characteristic temperature* of the substance, and may be calculated from the velocities of waves in the substance by means of the equations above, which give ψ in terms of ν_m, and ν_m in terms of v_1 and v_2. The velocities v_1 and v_2 can be calculated from the bulk and rigidity moduli of elasticity. It is found that the specific heats of solids calculated in this way agree very well with those observed. In particular, at very low temperatures the specific heat is found to be proportional to the cube of the absolute temperature, and so becomes negligible at temperatures near zero.

The free energy of the solid can easily be calculated. We have seen that the free energy of a set of $\mathcal{N}$ similar oscillators of frequency ν is given by

$$F = k\mathcal{N}\theta\log(1 - e^{-h\nu/k\theta}) + \mathcal{N}\epsilon_0.$$

The free energy of the solid is the sum of the free energies corresponding to all the sets of vibrations in it, so that

$$F = E_0 + 4\pi k\theta\left(\frac{2}{v_1^3} + \frac{1}{v_2^3}\right)\int_0^{\nu_m}\nu^2\log(1 - e^{-h\nu/k\theta})\,d\nu,$$

where E_0 is the energy in the solid when there is no vibration. As before, put $x = h\nu/k\theta$ and $h\nu_m = k\psi$, so that

$$F = E_0 + \frac{9\mathcal{N}k\theta^4}{\psi^3}\int_0^{\psi/\theta}x^2\log(1 - e^{-x})\,dx.$$

* For the method of evaluating the integral see section 11.

When θ is very small, the integral becomes $\int_0^\infty x^2 \log(1-e^{-x})\,dx = -\pi^4/45$, so that

$$F = E_0 - \frac{\pi^4 \mathcal{N} k\theta^4}{5\psi^3}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The entropy of the solid at low temperatures is then given by

$$\Phi = -\frac{\partial F}{\partial \theta} = \frac{4\pi^4 \mathcal{N} k\theta^3}{5\psi^3}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

According to this, when $\theta = 0$, then $\Phi = 0$, for any solid. But at $\theta = 0$ all substances are solids, so that $\Phi = 0$ at $\theta = 0$ for all substances. This result is *Nernst's Heat Theorem*, sometimes called the *Third Law of Thermodynamics.**

Since $\Phi = -\frac{\partial F}{\partial \theta}$, we have $\frac{\partial F}{\partial \theta} = 0$ at $\theta = 0$. The Gibbs-Helmholtz equation $E = F - \theta\frac{\partial F}{\partial \theta}$ shows that $E = F$ at $\theta = 0$, and also that $\frac{\partial E}{\partial \theta} = 0$ at $\theta = 0$, so that at $\theta = 0$, $\frac{\partial F}{\partial \theta}$ and $\frac{\partial E}{\partial \theta}$ are both zero.

11. Theory of Heat Radiation.

Instead of a solid cubical block we consider a hollow cube filled with radiation. The energy density of this radiation is independent of the nature of the walls and depends only on their temperature. If the walls are perfect reflectors, the radiation will form stationary trains of waves, and the number of possible frequencies between ν and $\nu + d\nu$ will be $\frac{8\pi\nu^2 d\nu}{c^3}$, where c is the velocity of light in a vacuum. This follows, by differentiation with respect to ν, from the expression $\frac{4}{3}\pi(\nu/v)^3$ found in section 10 for the number of frequencies between 0 and ν per unit volume. The additional factor 2 arises because the light waves are transverse, so that the radiation travelling in any direction may be regarded as made up of two parts polarized in two perpendicular directions, just as in the case of the transverse waves in the solid block previously considered. The energy in any one of the stationary vibrations in the box will be determined by the amplitude of the vibration just as in the case of the oscillators, so that according to the quantum theory the possible energies of the vibration will be given by $\epsilon = nh\nu$ where $n = 0, 1, 2, 3$, &c. The average energy of the vibrations having frequencies between ν and $\nu + d\nu$ will therefore be $\frac{h\nu}{e^{h\nu/k\theta} - 1}$, and the energy density $E_\nu d\nu$ in these vibrations will be given by

$$E_\nu = \frac{8\pi\nu^3 h}{c^3(e^{h\nu/k\theta} - 1)}.$$

* Cf. J. K. Roberts, *Heat and Thermodynamics*, Chap. XVIII.

Here E_ν may be defined as the energy density per unit range of frequency. This expression for E_ν was first obtained by Planck by means of the quantum theory. It is found to agree well with experiments on the distribution of the energy in the spectrum of black body radiation. When $h\nu/k\theta$ is large, the expression takes the form

$$E_\nu = \frac{8\pi\nu^3 h}{c^3} e^{-h\nu/k\theta},$$

known as Wien's formula. When $h\nu/k\theta$ is very small, $e^{h\nu/k\theta} = 1 + h\nu/k\theta$, and Planck's expression becomes

$$E_\nu = \frac{8\pi\nu^2 k\theta}{c^3},$$

so that each vibration has energy $k\theta$, as in the case of the oscillators and solid body According to Newtonian dynamics we should expect all values of the energy of a vibration or oscillator to be possible, since energy is supposed to be capable of continuous variation. The possible values of the energy of a vibration on the quantum theory are given by $\epsilon = nh\nu$, $n = 0, 1, 2, 3$, &c, so that if h were indefinitely small all values of the energy would be possible since n can be as large as we please. The part of the quantum theory so far considered differs from the classical theory only in the supposition that h has a finite value instead of an indefinitely small value. If we put $h = 0$ in Planck's formula for E_ν it becomes

$$E_\nu = \frac{8\pi\nu^2 k\theta}{c^3}.$$

This result is therefore that to which Newtonian dynamics leads, and it agrees with Planck's formula when $h\nu/k\theta$ is very small, that is for low frequencies and high temperatures. The total energy density E in the radiation is given by

$$E = \int_0^\infty E_\nu d\nu.$$

If we put $E_\nu = 8\pi\nu^2 k\theta/c^3$ we get an infinite value of E, which is of course impossible. The observed distribution of energy in the spectrum agrees with Planck's formula and differs entirely from that given by the classical theory even for frequencies of quite ordinary values. Newtonian dynamics therefore fails to explain heat radiation and so cannot be universally true as was formerly supposed. Of the other formulæ obtained in this chapter by means of the quantum theory, all those which do not contain h agree with Newtonian dynamics, while those containing h entirely disagree. The formulæ of the quantum theory all agree with experience, while those of the classical theory only do so when they agree with the quantum theory also.

If we put

$$E_\nu = \frac{8\pi\nu^3 h}{c^3(e^{h\nu/k\theta} - 1)} = \frac{8\pi k^3\theta^3}{c^3 h^2}\,\frac{x^3}{e^x - 1},$$

we get
$$E = \frac{8\pi h}{c^3}\int_0^\infty \frac{\nu^3 d\nu}{e^{h\nu/k\theta} - 1} = \frac{8\pi k^4\theta^4}{c^3 h^3}\int_0^\infty \frac{x^3 dx}{e^x - 1},$$

where $x = h\nu/k\theta$ Then

$$\int_0^\infty \frac{x^3 dx}{e^x - 1} = \int_0^\infty (e^{-x} + e^{-2x} + e^{-3x} + \ldots)\, x^3 dx$$

$$= 6\left(1 + \frac{1}{2^4} + \frac{1}{3^4} - \ldots\right) = \frac{\pi^4}{15}.$$

Hence
$$E = \frac{8}{15}\,\frac{\pi^5 k^4\theta^4}{c^3 h^3}.$$

Thus according to Planck's formula the total energy density of black body radiation in an enclosure is proportional to the fourth power of the absolute temperature, in agreement with Stefan's law. The relation between $y = \dfrac{x^3}{e^x - 1}$ and x is shown in fig. 2. We see that as x increases $y = \dfrac{x^3}{e^x - 1}$ increases to a maximum value and then diminishes. E_ν the energy density per unit range of frequency is equal to $8\pi k^3\theta^3 y/c^3h^2$, and so for any given temperature E_ν is proportional to y, and for any given value of y E_ν is proportional to the cube of the absolute temperature.

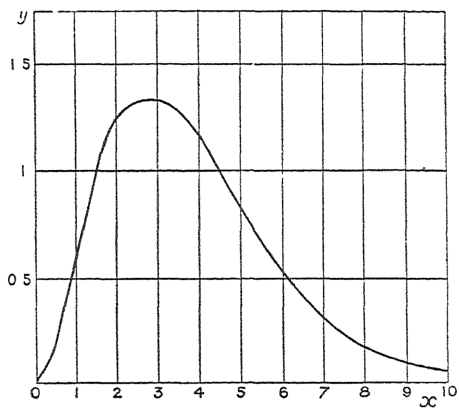


Fig 2

The energy density E_λ of the radiation per unit range of wave-length λ can be got by putting $\nu = c/\lambda$ in $E_\nu d\nu = \dfrac{8\pi h\nu^3 d\nu}{c^3(e^{h\nu/k\theta} - 1)}$, and writing $E_\nu d\nu = -E_\lambda d\lambda$. This gives

$$E_\lambda = \frac{8\pi hc}{\lambda^5(e^{hc/k\theta\lambda} - 1)}.$$

The wave-length λ_m for which E_λ is a maximum is got by putting $\frac{dE_\lambda}{d\lambda} = 0$. This gives

$$e^{hc/k\theta\lambda_m} = \frac{5}{5 - hc/k\theta\lambda_m},$$

the solution of which is $\frac{hc}{k\theta\lambda_m} = 4{\cdot}9651 \ldots$. The two equations $E = \frac{8\pi^5k^4\theta^4}{15c^3h^3}$ and $\frac{hc}{k\theta\lambda_m} = 4{\cdot}9651$ enable the constants k and h to be calculated from the experimental values of E and λ_m. In this way it is found that $h = 6{\cdot}53 \times 10^{-27}$ erg-sec., and $k = 1{\cdot}37 \times 10^{-16}$ erg/degree. The most probable values of these constants are believed to be $h = 6{\cdot}62 \times 10^{-27}$ and $k = 1{\cdot}379 \times 10^{-16}$. The energy $h\nu$ is usually called a quantum of energy of frequency ν. If ν_m denotes the frequency of the light for which E_λ is a maximum so that $c = \lambda_m\nu_m$ then we have $h\nu_m = 4{\cdot}9651\ k\theta$. The average kinetic energy of one molecule of a gas, as we have seen, is $\frac{3}{2}k\theta$, so that the quantum $h\nu_m$ is equal to 3·31 times the average kinetic energy of a gas molecule at the same temperature. The fraction of the possible vibrations of frequency ν which have energy $nh\nu$ is equal to $w_n = ae^{-nh\nu/k\theta}$ (cf. section 5; and $\frac{1}{a} = \Sigma e^{-nh\nu/k\theta} = 1/(1 - e^{-h\nu/k\theta})$, or $w_n = e^{-nh\nu/k\theta}(1 - e^{-h\nu/k\theta})$. The fraction which have no energy is therefore $w_0 = 1 - e^{h\nu/k\theta}$, so that the fraction which have any energy is $e^{-h\nu/k\theta}$. For example, at 0° C. the fraction of the vibrations of the frequency of yellow light ($\nu = 5 \times 10^{14}$) which have any energy is only about e^{-87} or 10^{-38}. The fraction of the vibrations of frequency ν_m which have any energy is $e^{-4\,965}$ or about one in 140. According to classical dynamics the average energy of the vibrations of any frequency should be $k\theta$.

12. Einstein's Theory of Heat Radiation.

So far we have supposed that the energy is distributed among the parts of the system considered in such a way as to make the entropy or the number of possible microscopic states a maximum, but we have not considered the nature of the actions between the parts by which this equilibrium state is produced. An interesting way (due to Einstein) of getting Planck's formula throws light on this question, and will now be considered. Let a large number of molecules or atoms be contained in a perfectly reflecting enclosure so that there must be equilibrium between the black body radiation in the enclosure and the atoms. Also, let each atom be only capable of existing in one or other of a series of states, having energies ε_1, ε_2, ε_3 . . ., and let the numbers of atoms in these states be N_1, N_2, N_3 . . . . Then the energy density of the radiation E_ν per unit range of frequency ν may be taken to be given by Wien's law,

$$E_\nu = \nu^3 f(\theta/\nu),$$

which can be deduced from purely thermodynamical considerations; and the

numbers of the atoms in the different states may be taken to be given by

$$N_1 = p_1 e^{-\epsilon_1/k\theta}, \ N_2 = p_2 e^{-\epsilon_2/k\theta} \dots .$$

The radiation of frequency ν is supposed to be absorbed by atoms in the state having energy ε_n, and as the result of this absorption these atoms are changed to another state having energy ε_m. The number of atoms changed from state n to state m in this way will be proportional to N_n and to E_ν, and so may be put equal to $\alpha N_n E_\nu$, where α is a constant. The reverse change is supposed also to occur and in the same way the number changing from m to n may be put equal to $\beta N_m E_\nu$. The change from m to n is also supposed to occur spontaneously without any action of the radiation, and the number changing in this way may be put equal to γN_m. In a state of equilibrium then we have

$$\alpha N_n E_\nu = \gamma N_m + \beta N_m E_\nu,$$

or

$$E_\nu (\alpha N_n - \beta N_m) = \gamma N_m.$$

Putting $N_n = p_n e^{-\epsilon_n/k\theta}$, and $N_m = p_m e^{-\epsilon_m/k\theta}$,

we get

$$E_\nu(\alpha p_n e^{(\epsilon_m - \epsilon_n)/k\theta} - \beta p_m) = \gamma p_m.$$

At high temperatures E_ν becomes very large, and γp_m may be supposed independent of the temperature, so that we must have $\alpha p_n - \beta p_m = 0$,

and therefore

$$E_\nu \alpha p_n (e^{(\epsilon_m - \epsilon_n)/k\theta} - 1) = \gamma p_m.$$

Now E_ν only contains θ in the function $f(\theta/\nu)$ of θ/ν, so that $(\varepsilon_m - \varepsilon_n)/k\theta$ must also be a function of θ/ν, which suggests that $\varepsilon_m - \varepsilon_n$ should be proportional to ν. Let $\varepsilon_m - \varepsilon_n = h\nu$, where h is a constant, so that

$$E_\nu = \frac{\gamma p_m/\alpha p_n}{e^{h\nu/k\theta} - 1}.$$

At high temperatures this gives

$$E_\nu = \frac{(\gamma p_m/\alpha p_n)k\theta}{h\nu}.$$

But at high temperatures the quantum theory and classical dynamics agree in giving

$$E_\nu = 8\pi\nu^2 k\theta/c^3,$$

so that we must have

$$\frac{(\gamma p_m/\alpha p_n)}{h\nu} = \frac{8\pi\nu^2}{c^3},$$

and therefore at any temperature

$$E_\nu = \frac{8\pi\nu^3 h}{c^3(e^{h\nu/k\theta} - 1)},$$

which is Planck's formula.

The radiation of frequency ν absorbed by the atoms in the above calculation must be equal to that emitted, so that when an atom goes from state m to state n we must suppose that it emits radiation of frequency ν having energy $\varepsilon_m - \varepsilon_n$. According to this, when an atom emits an amount of energy, say ε, in the form of radiation, the frequency of the radiation emitted is given by the equation $\varepsilon = h\nu$. This result is the basis of Bohr's theory of spectral lines and was first put forward by him.

13. Fermi-Dirac Theory of Electron Gas.

The theory of a monatomic gas given above (Section 7) is not applicable to electron gas, at high pressures, because according to Pauli's exclusion principle not more than one electron in a system can be in the same state. Consider electron gas with N electrons in a volume V. Let the momentum components of an electron be p_x, p_y, p_z, and consider the electrons with momenta between p and $p + dp$ for which the points with co-ordinates p_x, p_y, and p_z lie in a spherical shell of volume $4\pi p^2 dp$ in the momentum space.

Let there be Q_s possible states for the electrons in this shell, so that it may be regarded as divided into Q_s compartments each of which according to the exclusion principle can only contain either one electron or none. Let there be Z_{0s} of the Q_s compartments containing no electrons and Z_{1s} containing one, so that $Q_s = Z_{1s} + Z_{0s}$. The number of different ways W_s in which the electrons can be arranged in the different compartments is

$$W_s = \frac{Q_s!}{Z_{1s}!\, Z_{0s}!}.$$

For if all the different compartments contained different numbers of electrons, the Q_s groups of electrons could be arranged in $Q_s!$ different ways among the Q_s different compartments. But when Z_{1s} and Z_{0s} groups are all alike, so that it makes no difference how they are arranged, then W_s has the above value. Using the approximate value $\log N! = N \log N - N$, we get

$$\log W_s = Q_s \log Q_s - Z_{0s} \log Z_{0s} - Z_{1s} \log Z_{1s}.$$

All the electrons in the shell have energies $E_s = p^2/2m$, so the total energy in the shell is $Z_{1s}E_s$.

The whole volume of the momentum space may be supposed divided into concentric thin shells $Q_1, Q_2, Q_3 \ldots, Q_s \ldots$, so that if W denotes the total number of different arrangements, then

$$\log W = \sum_s \log W_s = \sum_s (Q_s \log Q_s - Z_{1s} \log Z_{1s} - Z_{0s} \log Z_{0s}).$$

Also
$$Q_s = Z_{0s} + Z_{1s},$$
$$N = \sum_s Z_{1s} \quad \text{and} \quad E = \sum_s E_s Z_{1s}$$

are constants. The equilibrium distribution of the electrons is obtained by making $\log W$ a maximum with Q_s, N and E constant. This gives

$$\sum_s (\delta Z_{1s} \log Z_{1s} + \delta Z_{1s} + \delta Z_{0s} \log Z_{0s} + \delta Z_{0s}) \\ + \alpha \sum_s (\delta Z_{0s} + \delta Z_{1s}) + \beta \sum_s E_s \delta Z_{1s} + \gamma \Sigma\, \delta Z_{1s} = 0,$$

where α, β, γ are constants. Equating to zero the coefficients of the variations, we get

$$1 + \log Z_{1s} + \alpha + \beta E_s + \gamma = 0,$$

$$1 + \log Z_{0s} + \alpha = 0,$$

so that

$$\log(Z_{0s}/Z_{1s}) = \gamma + \beta E_s$$

or

$$Z_{0s} = Z_{1s} e^{\gamma + \beta E_s}.$$

The number N_s of electrons in the shell s is therefore given by

$$N_s = Z_{1s} = Q_s \frac{Z_{1s}}{Z_{0s} + Z_{1s}} = Q_s \frac{1}{\frac{Z_{0s}}{Z_{1s}} + 1},$$

so that

$$N_s = Q_s \frac{1}{e^{\gamma + \beta E_s} + 1}.$$

The volume of a cell in the state space $\int dx\,dy\,dz\,dp_x\,dp_y\,dp_z$ is equal to h^3, so that the volume of a compartment in the momentum space is h^3/V since all the electrons are in the volume V, which is $\int dx\,dy\,dz$. Q_s is therefore equal to $4\pi V p^2 dp/h^3$, and, putting $\gamma = -\beta W$, we get

$$N_s = \frac{4\pi V p^2 dp}{h^3} \frac{1}{e^{\beta(E-W)} + 1}.$$

This is the number of electrons with momenta between p and $p + dp$.

The constant β may be shown to be equal to $1/k\theta$ as in Planck's theory of entropy.*

The number of electrons per cm.^{3} with momentum co-ordinates between p_x and $p_x + \delta p_x$, p_y and $p_y + \delta p_y$, and p_z and $p_z + \delta p_z$ is therefore given by

$$f \delta p_x \delta p_y \delta p_z = \frac{\delta p_x \delta p_y \delta p_z}{h^3(e^{(E-W)/k\theta} + 1)}.$$

When $\theta = 0$, or very small, this makes f equal to zero when $E > W$, and equal to $1/h^3$ when $E < W$. According to this, at very low temperatures, each compartment of volume h^3 for which the energy E, or $p^2/2m$, is less than W contains one electron, while the compartments for which the energy is greater than W are all empty.

So far we have not taken into account the spin of the electrons. According to the spinning electron theory,† an electron has two possible states for any possible momentum value, so that there are

* See p. 74. † See p. 192.

two practically independent sets of electrons. This just doubles the value of f, so that

$$f = \frac{2}{h^3(e^{(E-W)/k\theta}+1)}.$$

At very low temperatures, then, all the cells contain two electrons with opposite spins for energy values up to W. If $W = p_m{}^2/2m$, then p_m is the greatest value of p when $\theta = 0$, so that

$$\tfrac{4}{3}\pi p_m{}^3 = \tfrac{1}{2}Nh^3,$$

where N is the number of electrons in one cm.^{3}, or

$$W = \frac{p_m{}^2}{2m} = \frac{h^2}{2m}\left(\frac{3N}{8\pi}\right)^{2/3}.$$

The value of W is therefore proportional to $N^{2/3}$. If we put $W = \phi e$, then ϕ is the potential difference required to give an electron energy equal to W and $\phi = \dfrac{h^2}{2me}\left(\dfrac{3N}{8\pi}\right)^{2/3}$. This expression gives the following values of ϕ in volts.

N (Electrons per cm.^{3})	ϕ (Volts)
10^{21}	0 38
10^{22}	1·76
10^{23}	8·20
10^{24}	38
10^{25}	176

It appears that $W = \phi e$ is small unless the number of electrons per cm.^{3} is very large. In an electron gas at, say, 1 mm. of mercury pressure N would be about $3{\cdot}5 \times 10^{16}$, so that W would be quite negligible and we should have $f = \dfrac{2}{h^3(e^{E/k\theta}+1)}$. This leads to the Maxwell distribution of velocities approximately when $E/k\theta$ is large. The number of free electrons in metals is of the order of magnitude 10^{23} per cm.^{3}, so that W is quite large, and at temperatures even up to 1000° C. the velocity distribution differs little from that at the absolute zero. The energy of the electrons, therefore, is practically independent of the temperature. Thus it appears that the Maxwell distribution of velocities holds approximately for an electron gas at low pressures, but for the free electrons in metals it does not hold and must be replaced by the Fermi-Dirac distribution, which makes the energy of the electrons practically independent of the temperature.

14. Bohr's Theory of the Hydrogen Atom.

Bohr supposed that an atom can exist only in a series of states having energies E_1, E_2, E_3, . . ., and that when it changes from a state

having energy E_n to another having less energy E_m the frequency ν of the radiation emitted is given by $E_n - E_m = h\nu$.

A hydrogen atom consists of a nucleus with charge e and an electron moving round the nucleus. The mass of the nucleus is about 1840 times that of the electron, so that the nucleus may be supposed to be at rest. If the electron is moving with velocity v in a circular orbit of radius r, then we have $mv^2/r = e^2/r^2$, so that the kinetic energy $\frac{1}{2}mv^2$ is equal to $e^2/2r$ and the total energy E is equal to $-e^2/2r$ since the potential energy is $-e^2/r$.

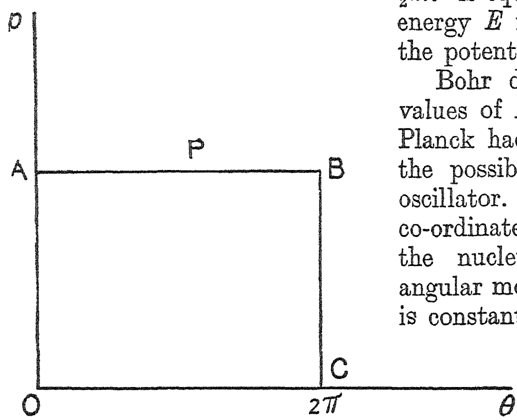


Fig. 3

Bohr determined the possible values of E in the same way that Planck had previously determined the possible energies of a simple oscillator. If r and θ are the polar co-ordinates of the electron, with the nucleus at the origin, the angular momentum of the electron is constant and equal to $p = mr^2\dot{\theta}$. If we take p and θ for the co-ordinates of a point P in fig. 3, then as θ increases from 0 to 2π the point P moves from A to B. Bohr supposed that the area of the rectangle ABCO must be equal to a multiple of h or that $\int_0^{2\pi} p\, d\theta = nh$. In the case of the simple oscillator with equation of motion $m\ddot{x} = -\mu x$, if we put $y = m\dot{x}$ and take x and y for the co-ordinates of a point, then the point moves round an ellipse, and Planck supposed that the area of the ellipse must be a multiple of h as we have seen.

Since p is constant the equation $\int_0^{2\pi} p\, d\theta = nh$ gives $p = nh/2\pi$. In the case of a circular orbit $p = mvr$, so that $mv = nh/2\pi r$. The kinetic energy $(mv)^2/2m = n^2h^2/4\pi^2r^2 2m = e^2/2r$, so that $r = n^2h^2/4\pi^2me^2$, and the total energy $E = -e^2/2r$ is given by $E_n = -2\pi^2me^4/n^2h^2$.

If the atom changes from a state with energy E_n to another with energy $E_{n'}$, then Bohr supposed that the energy difference $E_n - E_{n'}$ is emitted as radiation of frequency ν given by $h\nu = E_n - E_{n'}$, so that

$$h\nu = \frac{2\pi^2me^4}{h^2}\left\{\frac{1}{n'^2} - \frac{1}{n^2}\right\},$$

where n must be greater than n'. If we take $n' = 1$, we get a series of frequencies with $n = 2, 3, 4, \ldots$, and with $n' = 2$ another series with $n = 3, 4, 5, \ldots$, and so on.

The observed spectrum of atomic hydrogen contains just such series of lines. The Balmer series has frequencies given very exactly by $\nu = 3{\cdot}290 \times 10^{15}\left(\frac{1}{4} - \frac{1}{n^2}\right)$ with $n = 3, 4, 5, \ldots$. The Lyman series is given by $\nu = 3{\cdot}290 \times 10^{15}\left(1 - \frac{1}{n^2}\right)$, and the Paschen series by $\nu = 3{\cdot}290 \times 10^{15}\left(\frac{1}{9} - \frac{1}{n^2}\right)$.

According to Bohr's theory, the constant $3{\cdot}290 \times 10^{15}$ should be equal to $2\pi^2 me^4/h^3$, and on substituting the known values * of m, e and h we find $2\pi^2 me^4/h^3 = 3{\cdot}294 \times 10^{15}$, which agrees with the value got from the observed frequencies within the limits of error.

If the electron is moving in an elliptical orbit, then it has radial as well as angular momentum. Bohr supposed that the possible values of the radial momentum $m\dot{r}$ are given by $\int m\dot{r}\,dr = nh$, where n is an integer and the integral is taken once round the orbit. If $\int p\,d\theta = n_1 h$ and $\int m\dot{r}\,dr = n_2 h$, then it is found that $E_n = -2\pi^2 me^4/h^2(n_1 + n_2)^2$ and that the ratio of the minor to the major axis of the ellipse is equal to $n_1/(n_1 + n_2)$.

Bohr's theory was also applied to atoms containing more than one electron, but the possible energies of such atoms cannot be accurately calculated because the motion of two or more electrons about a nucleus cannot be solved exactly. Nevertheless it enabled many facts about the spectra of such atoms to be explained. The equations $E_n - E_{n'} = h\nu$ and $\int p\,dq = nh$ were assumed in the quantum theory, and the assumption was justified by the success of the theory in explaining the facts of heat radiation, specific heats and spectra. This theory was unsatisfactory because of the arbitrary nature of the assumptions made, and also because it failed to account for many of the finer details of spectra and other phenomena such as absorption and dispersion.

15. Quantum Mechanics.

To meet these difficulties new ideas were introduced and new methods developed by de Broglie in 1924 and by Heisenberg, Schrödinger and Dirac in 1925 and 1926. The new methods developed by these authors and many others have replaced the old quantum theory of Planck and Bohr. The new quantum theory is known as *quantum mechanics.*

The photo-electric effect shows, as we have seen, that the effects produced by light or X-rays of frequency ν are such as might be expected from particles of energy $h\nu$. The distribution of these effects, however, is such as would be expected for a distribution of wave

* See p. 434

intensities. The phenomena of interference and diffraction of light and X-rays agree perfectly with the wave theory, but the effects produced require us to assume that the effects are produced by particles called photons with energy $h\nu$. The number of particles falling on a unit area, or rather the number of effects observed which are such as might be expected from particles, is proportional to the wave intensity.

To make this clear consider the simple optical experiment shown in fig. 4. At S is a straight electric-light filament perpendicular to the plane of the figure. ABC is a glass prism with the angle ABC nearly equal to 180°, and DE is a white screen. The prism deviates the light falling on it between A and B downwards and that between B and C upwards, so that the two beams are superposed on the screen between

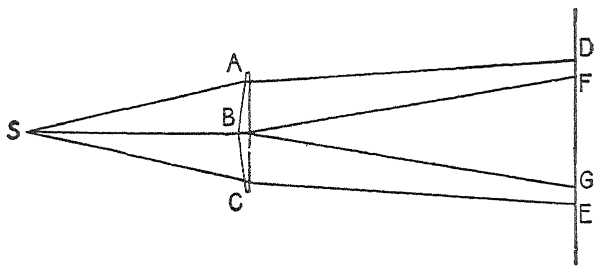


FIG. 4

F and G. We then get bright and dark interference bands on the screen between F and G as is well known. This result is explained perfectly by the wave theory, and the wave-length of the light can be calculated from the distance between the bright bands. The bands can be photographed if the screen is replaced by a photographic plate. If the plate is only exposed for a very short time, then it is found that only a few of the grains of silver bromide in it are affected, the rest are unchanged. The affected grains which show up when the plate is developed are distributed in bands like the light on the screen. On the wave theory we should expect all the grains in a bright band to be affected. The light acts like particles in this experiment, so that only the grains hit by a particle are affected, but the particles are distributed like waves. If instead of a photographic plate we use a photo-electric cell, then we find that electrons are shot out of the atoms in the bright bands. The electrons shot out receive energy $h\nu$ from the light, but only a very small fraction of the electrons are affected. On the wave theory we should expect all the electrons to get a small amount of energy, whereas actually a few get the relatively large amount $h\nu$. The light acts like particles with energy $h\nu$, but the particles are distributed like a wave intensity.

It appears that in any optical problem we can calculate the effects to be expected by assuming a wave theory and calculating the wave intensity. The effects to be expected are then such as would be produced by particles of energy $h\nu$, and the number of the effects is proportional to the calculated wave intensity, or we may say that the chance of an effect occurring on any small area in a given time is proportional to the area and to the wave intensity at the area. With very weak light the number of effects will be small, so that the chance of getting an effect on a particular small area may be very small unless the time interval is very long.

It has been shown experimentally that a diffraction pattern photographed with strong light and a very short exposure is identical with one produced in the same way with very weak light and a very long exposure. This is what we should expect if the distribution of the effects is determined by the relative wave intensities, so that the fraction of the effects occurring at any place is independent of the intensity of the source. If we suppose that the source of light emits only one photon, the chance of an effect due to this photon occurring at any place will be proportional to the wave intensity, at the place, when the source is supposed to be emitting a continuous train of waves. The waves therefore carry no energy, and the wave theory may be regarded as merely auxiliary mathematics which enables the distribution of the photons to be calculated. Just why such a method of calculation is necessary and why it gives results in agreement with the facts is not known.

In 1924 de Broglie suggested that the distribution of particles like electrons should also be calculated by a wave theory. The energy of a photon is equal to $h\nu$, so that its mass m is given by $h\nu = mc^2$, where c is the velocity of light. The velocity of a photon is always equal to c, so its momentum is equal to mc, which is equal to $h\nu/c$. The wave-length λ of the waves associated with the photon is equal to c/ν, so that its momentum is equal to h/λ. De Broglie supposed that similar relations hold for electrons and the waves associated with them. If m denotes the mass of an electron moving with velocity v, then its momentum is mv, so that $mv = h/\lambda$ or $\lambda = h/mv$. Also its energy $mc^2 = h\nu$. The velocity of the electron waves u is equal to $\nu\lambda$ and so is equal to $(h/mv) \times (mc^2/h)$, so that $uv = c^2$. Since v is less than c, the wave velocity u is greater than c. If $v = c/10$, then $u = 10c$. The wave-length of the waves associated with an electron which has been set in motion by a potential difference V may be calculated as follows. The mass m is equal to $m_0 + Ve/c^2$, where m_0 is the mass of an electron at rest. The velocity v of the electron is given by

$$Ve = m_0c^2((1 - v^2/c^2)^{-1/2} - 1).$$

Substituting in $\lambda = h/mv$, we get $\lambda = hc/\sqrt{Ve(2m_0c^2 + Ve)}$. This

equation, when $Ve/2m_0c^2$ is small, gives $10^8\lambda = \sqrt{150/V}$ with V in volts. It gives the following values of λ:

V (Volts)	λ (Cm $\times 10^{-8}$)
1	12 25
10	3 87
100	1·225
1000	0 387

Thus the de Broglie waves of an electron are of the same order of length as X-rays.

De Broglie's suggestion that waves are associated with particles like electrons has been confirmed experimentally by Davisson and Germer and later by G. P. Thomson and others.

Davisson and Germer allowed a narrow beam of electrons, all moving with nearly the same velocity, to fall on a face of a crystal of nickel in a high vacuum. They found that the electrons were reflected from the crystal face just as X-rays are reflected. When X-rays of wave-length λ fall on a crystal face at a glancing angle θ, then they are strongly reflected if $n\lambda = 2d \sin\theta$, where d is the distance between the layers of atoms parallel to the crystal face. If $n\lambda$ is not equal to $2d \sin\theta$, then there is no appreciable reflection. Davisson and Germer found that the same thing is true for the electron beam with λ equal to the wave-length of the de Broglie waves of the electrons in the crystal.

Davisson and Germer used electrons having velocities due to potential differences from about 50 to 500 volts. G. P. Thomson has confirmed their results with electrons having velocities due to much higher potentials around 15,000 volts. In his experiments the electrons were passed through extremely thin films of gold and other metals and received on a photographic plate. The gold consists of many minute crystals orientated at random. The minute crystals diffract the electron waves just as X-rays are diffracted when passed through a finely powdered crystalline substance. Concentric rings are formed on the plate, and the angles of diffraction can be found from the diameters of the rings. G. P. Thomson found that the wave-lengths deduced from the diffraction angles agreed with de Broglie's formula $\lambda = h/mv$.

It appears that electrons behave like trains of waves and also like particles just as light does. In each case the frequency of the waves is given by $E = h\nu$, where E is the energy of the particle and the wave-length is equal to Planck's constant h divided by the momentum of the particle.

16. Ray Paths and Particle Paths.

In geometrical optics rays of light are considered, and these rays travel in straight paths, in a uniform medium, and are reflected and refracted according to well-known laws. The path of a ray from a point A to another point B, when the refractive index of the medium between A and B varies from point to point in any way, may be found by means of Fermat's law, according to which the time for the light to go from A to B along its actual path is equal to the time along any other path from A to B which lies very near to the actual path. If ds is an element of a path from A to B and u the velocity of the light along ds, then Fermat's law is equivalent to $\delta\int_A^B ds/u = 0$ for the actual path. $\int_A^B ds/u$ is the time from A to B, and $\delta\int_A^B ds/u$ is supposed to indicate the change of the integral due to a change from the actual path to another path very close to it. Fermat's law gives the ray path for waves of any kind, and so should give the path of an electron or other particle provided u is taken to be the wave velocity of the electron waves. If the velocity of the electron is v, then $u = c^2/v$, so that Fermat's law gives $\delta\int_A^B v\,ds = 0$, or $\delta\int_A^B mv\,ds = 0$, where m is the mass of the particle. $\int_A^B mv\,ds$ is called the action, and $\delta\int_A^B mv\,ds = 0$ expresses the principle of least action, which, as is well known, gives the path of a particle moving in a field of force. In this way we see that the path of a particle as determined by classical dynamics agrees with the ray path of the waves associated with the particle according to de Broglie's theory.

Geometrical optics or the theory of ray paths is valid only when the dimensions of the cross-section of the ray are large compared with the wave-length of the waves. Ordinary light has wave-lengths about 10^{-4} cm., so that if we pass light from a small source through a hole one millimetre in diameter we get a ray which will go a long way without deviating seriously from the laws of geometrical optics, but if we pass the light through a hole 10^{-4} cm. in diameter we do not get a ray because the light diverges in all directions from the hole. The laws of geometrical optics are approximately true for large-scale phenomena with dimensions very large compared with the wave-length. For small-scale phenomena the wave theory must be used, and the light intensity at any point must be calculated by the methods used in the theory of interference and diffraction.

The path of a particle calculated by classical dynamics agrees with the ray path of the de Broglie waves and so presumably is an approximation valid only for phenomena on a scale large compared with the wave-lengths. The de Broglie waves of an electron have

lengths around 10^{-8} cm. which is of the same order as the dimensions of an atom. We should therefore not expect classical dynamics to be useful for atomic phenomena. For such small-scale phenomena the wave intensities must be calculated as in the theory of interference and diffraction of light. Planck and Bohr obtained the possible energies of an electron describing an orbit by supposing that $\int p\,dq = nh$, where q is a co-ordinate of the electron and p the associated momentum. The integral is taken once round the orbit. If we take q to be the distance along the orbit and p the momentum along the direction of motion, we have for the wave-length of the de Broglie waves going round the orbit $\lambda = h/p$, so that $\int h\,dq/\lambda = nh$ or $\int dq/\lambda = n$. $\int dq/\lambda$ is the number of wave-lengths in the length of the orbit, so that the possible orbits are those divisible into parts each equal to a whole wave-length. If the train of waves extends many times round the orbit, the waves would interfere and destroy each other if the orbit could not be divided exactly into whole waves. Thus de Broglie's waves offer an explanation of the condition that $\int p\,dq = nh$. We see that this equation depends on the idea of a ray path coinciding with the orbit. It is therefore not necessarily exact and must be replaced by calculations based on wave theory.

According to de Broglie's ideas, then, the chance of finding a particle anywhere is proportional to the intensity or the square of the amplitude of the waves associated with the particle, just as in optics the chance of getting an effect due to a photon is proportional to the intensity of the light waves. If w denotes the wave amplitude, then $w^2 dv$ or $\bar{w}w\,dv$ when w is complex is proportional to the chance of finding the particle in an element of volume dv. If the values of w are so chosen that $\int \bar{w}w\,dv = 1$, the integration being taken over all regions in which w is not zero, then $\bar{w}w\,dv$ is equal to the chance of finding the particle in dv. w is then said to be normalized.

17. Wave Groups.

In classical dynamics the position of a free particle moving along the x-axis with a velocity v is given by the equation $x = vt + c$. In quantum mechanics the chance of finding the particle between x and $x + dx$ is $\bar{w}w\,dx$, so the moving particle requires that $\bar{w}w$ be zero except inside a small range Δx of x between $vt + c - \frac{1}{2}\Delta x$ and $vt + c + \frac{1}{2}\Delta x$. The region in which w is not zero then moves along with the velocity v. If Δx is very small, then the particle must be moving with a velocity nearly equal to v. This means that to represent the moving particle we require a group of de Broglie waves moving along with the velocity of the particle. The wave velocity u is not

equal to the particle velocity v, but the velocity of a group of waves is not equal to the wave velocity when the wave velocity varies with the frequency. Suppose two trains of waves with slightly different frequencies ν and ν' and wave-lengths λ and λ' superposed. Let $n = 1/\lambda$ and $n' = 1/\lambda'$ so that n is the wave number. Then

$$w = A \sin 2\pi(\nu t - nx) + A \sin 2\pi(\nu' t - n'x)$$

when the amplitude A is the same for both trains. This gives

$$w = 2A \sin 2\pi \left(\frac{\nu + \nu'}{2} t - \frac{n + n'}{2} x\right) \cos 2\pi \left(\frac{\nu - \nu'}{2} t - \frac{n - n'}{2} x\right).$$

Since ν and ν' and n and n' are nearly equal, this gives, putting

$$\nu - \nu' = \delta\nu \quad \text{and} \quad n - n' = \delta n,$$

$$w = 2A \sin 2\pi(\nu t - nx) \cos 2\pi \left(\frac{\delta\nu t}{2} - \frac{\delta n x}{2}\right).$$

We have therefore a train of waves of frequency ν and wave number n with amplitude $2A \cos 2\pi \left(\frac{\delta\nu t}{2} - \frac{\delta n x}{2}\right)$. The two trains therefore give a series of wave groups. The group length l is given by $2\pi \frac{\delta n l}{2} = \pi$, so that $l\delta n = 1$. The amplitude is constant provided $\frac{\delta\nu t}{2} - \frac{\delta n x}{2}$ is constant, so that $\delta\nu t - \delta n x = c$. This gives $x = -\frac{c}{\delta n} + \frac{\delta\nu t}{\delta n}$, which shows that the points at which the amplitude is constant move along with the velocity $\delta\nu/\delta n$, so that this is the group velocity. For the de Broglie waves of a particle we have $\frac{1}{2}mv^2 = E - V$, so that $mv = \sqrt{2m(E - V)}$. Also $\lambda = 1/n = h/mv$, so that $mv = hn$, which with $E = h\nu$ gives $hn = \sqrt{2m(h\nu - V)}$ or $\nu = hn^2/2m + V/h$. This gives for the group velocity $\frac{d\nu}{dn} = hn/m$, which is equal to the particle velocity v. Thus it appears that a group of w waves does travel with the velocity of the particle associated with it. This was shown to be the case by de Broglie when he first suggested that waves are associated with a particle.

To form a group of waves it is necessary to superpose trains of waves with different frequencies, and the range of frequencies necessary depends on the length of the group. Suppose we have a group in which $w = Ae^{2\pi i n_0 x}$ at the time $t = 0$, between $x = -l/2$ and $x = +l/2$, but is equal to zero for values of x outside this range. The chance of finding the particle between x and $x + dx$ is $\bar{w}w\,dx$, which is equal to $A^2 dx$ inside the group between $x = -l/2$ and $x = +l/2$

but is zero outside. If the particle is supposed to be equally likely to be anywhere in the group, the chance should be dx/l, so that $A^2 = 1/l$. The amplitude $C(n)$ of the infinite wave trains required to produce this group is given by Fourier's integral theorem.* We have $w = \int_{-\infty}^{+\infty} C(n) e^{2\pi i n x}\, dn$, where $C(n) = \int_{-\infty}^{+\infty} w e^{-2\pi i n x}\, dx$. This gives

$$C(n) = \frac{1}{\sqrt{l}} \int_{-l/2}^{+l/2} e^{2\pi i (n_0 - n)x}\, dx,$$

so that

$$C(n) = \frac{1}{\sqrt{l}} \left[\frac{e^{\pi i (n_0 - n)l} - e^{-\pi i (n_0 - n)l}}{2\pi i (n_0 - n)} \right],$$

which gives

$$C(n) = \sqrt{l} \sin \pi (n_0 - n) l / \pi (n_0 - n) l.$$

Let $\alpha = \pi(n_0 - n)l$ so that $C(n) = \sqrt{l} \sin \alpha / \alpha$. $C(n)$ is zero when $\alpha = \pm\pi, \pm 2\pi, \pm 3\pi, \ldots$, and has maxima given by $\tan \alpha = \alpha$ or approximately when $\alpha = 0, \pm 3\pi/2, \pm 5\pi/2, \ldots$. The maximum values are equal to $\sqrt{l}, \sqrt{l}\frac{2}{3\pi}, \sqrt{l}\frac{2}{5\pi}, \ldots$. The intensity of the waves is proportional to the square of the amplitude, so that the intensities at the maxima are as $l, 4l/9\pi^2, 4l/25\pi^2, \ldots$. Thus nearly all the wave numbers with appreciable amplitudes lie between the values given by $\alpha = \pm\pi/2$. We may therefore take the wave numbers in the group to lie between the values given by $\pi(n_0 - n)l = \pi/2$ and $\pi(n_0 - n)l = -\pi/2$, so that if Δn denotes the range of wave numbers then $\Delta n l = 1$. Now $n = mv/h$, so that $\Delta n = \Delta(mv)/h$ and therefore $\Delta(mv)l = h$. $\Delta(mv)$ is the range of momentum values corresponding to the range of wave numbers in the group and so may be said to be the uncertainty in the momentum. The uncertainty in the position of the particle is equal to l since the particle must be somewhere in the group. The product of the two uncertainties is equal to h. It appears that the product of the uncertainties is of the order of magnitude of h. This is *Heisenberg's uncertainty principle.* If q is a co-ordinate and p the associated momentum, then $\Delta q \Delta p = h$ roughly, where Δq and Δp are the uncertainties in the observed values of q and p. Since $h = 6{\cdot}5 \times 10^{-27}$ the product is very small, much smaller than the product of the experimental errors in actual determinations of q and p. According to the uncertainty principle, it is impossible to know both q and p exactly. If q is known exactly so that $\Delta q = 0$, then Δp becomes infinite.

The group length l does not remain constant because the group velocity varies with the wave number, so that there is a range of group

* See p. 417.

velocities in the group corresponding to the range of wave numbers. The front of the group advances with the greatest group velocity and the rear with the least, so that the length of the group increases at a rate equal to the difference between the greatest and least group velocities. We have $\Delta(mv)l = h$, so that $\Delta v = h/lm$ is the rate of increase of the group length. As an example, consider an electron moving with a velocity of 10^9 cm./sec. and suppose its wave group is one millimetre long. In this case

$$h/lm = 6{\cdot}5 \times 10^{-27}/0{\cdot}1 \times 9 \times 10^{-28} = 70 \text{ cm./sec.}$$

The length of the group would therefore increase by $70 \times 10^{-7} = 7 \times 10^{-6}$ cm. while the electron moved 100 cm.

This spreading of the group increases the uncertainty in the position of the particle, so that the product $\Delta p \Delta q$ gets greater than h. The value h is therefore the least possible value of $\Delta p \Delta q$, and $\Delta p \Delta q$ may be much greater than h but cannot be much less.

The result that $\Delta p \Delta q$ is not less than about the value h may be obtained by considering experimental methods of measuring q and p. As an example, suppose we observe a particle in a microscope and try to estimate its position and momentum. The smallest distance that can be seen in a good microscope is about equal to the wave-length λ of the light used, so that the uncertainty in the position of the particle must be λ at least. When a photon is reflected by the particle, the momentum of the particle will be altered by an amount of the order of magnitude of the momentum of the photon or h/λ. The momentum of the particle will therefore be uncertain by at least h/λ, so that the product of the two uncertainties must be at least $\lambda(h/\lambda)$ or h.

The uncertainty in the energy E is equal to $p\Delta p/m$, and the uncertainty Δt in the time t at which the particle arrives at a point may be put equal to the group length l divided by v. Hence

$$\Delta E \,.\, \Delta t = p\Delta pl/mv = \Delta p \,.\, l = h.$$

Also $l \,.\, \Delta n = 1$ or $(l/v)v\Delta n = 1$, so that $(l/v)(\partial \nu/\partial n)\Delta n = 1$ and therefore $(l/v)\Delta \nu = 1$. But $E = h\nu$, so that $\Delta \nu = \Delta E/h$. Hence $(l/v)\Delta E = h$ or $\Delta t \Delta E = h$ as before.

18. **Schrödinger's Equation.**

Schrödinger first showed how to use de Broglie's ideas to obtain exact solutions of atomic problems. The differential equation, satisfied by a wave motion in three dimensions, is

$$\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} = \frac{1}{u^2}\frac{\partial^2 w}{\partial t^2} \quad \text{or} \quad \Delta w = \frac{1}{u^2}\frac{\partial^2 w}{\partial t^2},$$

where w is the wave amplitude and u the wave velocity. Schrödinger supposed that the waves of an electron must satisfy this equation. If the energy of the electron is E, then the equation $E = h\nu$ gives the frequency ν of the waves. We may therefore put $w = w_0 e^{-2\pi i\nu t}$, where w_0 is a function of x, y, z. This gives

$$\Delta w_0 = \frac{1}{u^2}(-2\pi i\nu)^2 w_0,$$

and since $\nu/u = 1/\lambda$ and $\lambda = h/mv$, we get

$$4\pi^2\nu^2/u^2 = 4\pi^2/\lambda^2 = 4\pi^2 m^2 v^2/h^2,$$

so that $4\pi^2\nu^2/u^2 = 8\pi^2 m(E - V)/h^2$ since $mv = \sqrt{2m(E - V)}$, where V is the potential energy, and so $\Delta w_0 + \frac{8\pi^2 m}{h^2}(E - V)w_0 = 0$, which is often called Schrödinger's equation.

Consider the case of a particle moving along the x-axis with constant velocity v and suppose that at $x = 0$ and $x = l$ there are plane surfaces which reverse the motion of the particle so that it moves backwards and forwards between $x = 0$ and $x = l$. In this simple case $V = 0$ and Schrödinger's equation becomes

$$\partial^2 w_0/\partial x^2 + 8\pi^2 mEw_0/h^2 = 0.$$

A solution of this is $w_0 = A \sin \alpha x$, which gives $-\alpha^2 + 8\pi^2 mE/h^2 = 0$, so that $\alpha^2 = 8\pi^2 mE/h^2$.

At $x = 0$ and $x = l$ we may suppose that $w_0 = 0$, so that $\sin \alpha l = 0$ and therefore $\alpha l = n\pi$ and $\alpha^2 = n^2\pi^2/l^2$, so that $E = n^2h^2/8ml^2$, where $n = 1, 2, 3, \ldots$.

Schrödinger supposed that the chance of finding the particle in between x and $x + dx$ would be proportional to $\bar{w}w\,dx$, so that with $w = A \sin \alpha x \,.\, e^{-2\pi i\nu t}$ the chance is $A^2 \sin^2 \alpha x\,dx$ and so is equal to zero when $\alpha x = N\pi$, where $N = 0, 1, 2, 3, \ldots$. Thus the chance is zero at $x = 0$ and $x = l$ and at points in between which divide l into equal parts. Half-way between these points the chance is a maximum. Since the particle must be somewhere between $x = 0$ and $x = l$, if we choose A so that $\int_0^l \bar{w}w\,dx = 1$, then $\bar{w}w\,dx$ will be equal to the chance of finding the particle in dx. This gives $\int_0^l A^2 \sin^2 \alpha x\,dx = 1$, so that $A^2 l/2 = 1$ or $A = \sqrt{2/l}$ and $w = \sqrt{2/l}\sin \alpha x \,.\, e^{-2\pi i\nu t}$. With this value of A, w is said to be normalized. This theory does not give the motion of the particle but only the chance of finding it in different positions. The solution obtained is quite different from that given by classical theory. On classical theory the energy E could have any

positive value, whereas we have found that $E = n^2h^2/8ml^2$ where $n = 0, 1, 2, 3, \ldots$. Also in a classical theory the chance of finding the particle between x and $x + dx$ would be dx/l, which is quite different from $\frac{2}{l}\sin^2 ax\,dx$.

This simple case may be compared with the transverse vibrations of a string fixed at both ends. If y denotes the transverse displacement, then $d^2y/dx^2 = (1/u^2)(d^2y/dt^2)$, where u is the wave velocity. Putting $y = y_0e^{-2\pi\nu t}$, we get $d^2y_0/dx^2 + 4\pi^2\nu^2y_0/u^2 = 0$ or $d^2y_0/dx^2 + 4\pi^2y_0/\lambda^2 = 0$. A solution is $y_0 = A\sin ax$, which gives $a = 2\pi/\lambda$ Since $y_0 = 0$ at $x = 0$ and $x = l$, we must have $\sin al = 0$ or $al = n\pi$, so that $2\pi l/\lambda = n\pi$ or $\lambda = 2l/n$, where $n = 1, 2, 3, \ldots$. We have stationary waves in the string with nodes where $y = 0$ at equally spaced points. The possible frequencies of vibration are given by $\nu = u/\lambda$, so that $\nu = un/2l$. The energy of the vibrating string may have any value but only certain frequencies are possible. In Schrödinger's theory the energy E is equal to $h\nu$, so that since only certain frequencies are possible only the corresponding energies are possible. The amplitude of vibration is not related to the energy but may be given any convenient value, for example the value found by normalizing w. The string may vibrate simultaneously with any of its possible frequencies but, if the particle is supposed to have a definite energy, the waves associated with it can only have one frequency.

19. Simple Oscillator.

Let us now consider the case of a particle moving along x and acted on by a force directed towards the origin and proportional to its distance x from the origin. We have then $md^2x/dt^2 = -\mu x$, where m is the mass of the particle and μ a constant. A solution is $x = A\sin at$ with $a = \sqrt{\mu/m}$. The time of one oscillation is given by $T = 2\pi\sqrt{m/\mu}$. This is the classical solution. On Schrödinger's theory we have, putting $V = \frac{1}{2}\mu x^2$, the equation

$$\frac{\partial^2 w_0}{\partial x^2} + \frac{8\pi^2 m}{h^2}(E - \tfrac{1}{2}\mu x^2)w_0 = 0,$$

where $w = w_0e^{-2\pi iEt/h}$. Schrödinger supposed that only finite solutions are admissible because an infinite value of w at a point would mean that the particle was certain to be at the point. Let $w_0 = Ue^{-ax^2}$, where U is a function of x only and a a constant. Substituting this value of w_0, we find

$$\frac{\partial^2 U}{\partial x^2} - 4ax\frac{\partial U}{\partial x} + \left\{\frac{8\pi^2 m}{h^2}(E - \tfrac{1}{2}\mu x^2) - 2a(1 - 2ax^2)\right\}U = 0.$$

The two terms in x^2 can be got rid of by taking $a^2 = \mu m\pi^2/h^2$, so that we have

$$\frac{\partial^2 U}{\partial x^2} - 4ax\frac{\partial U}{\partial x} + \left\{\frac{8\pi^2 m}{h^2}E - 2a\right\}U = 0.$$

If we assume a series of even powers of x, $a_0 + a_2x^2 + a_4x^4 + \ldots$ for U and substitute it in this equation, we find

$$(n+2)(n+1)a_{n+2} - 4ana_n + \left(\frac{8\pi^2 mE}{h^2} - 2a\right)a_n = 0.$$

This shows that a_{n+2} will be equal to 0, and the series will terminate, if the coefficient of a_n is zero or if $8\pi^2 mE/h^2 - 2a - 4an = 0$. This gives

$$E_n = \frac{h^2 a}{2\pi^2 m}(n + \tfrac{1}{2}),$$

or since $a = \pi\sqrt{\mu m}/h$, we get

$$E_n = \frac{h}{2\pi}\sqrt{\frac{\mu}{m}}(n + \tfrac{1}{2}),$$

where $n = 0, 2, 4, 6, \ldots$. In the same way if we take a series of odd powers $a_1x + a_3x^3 + \ldots$, we find that the series terminates if $E_n = \frac{h}{2\pi}\sqrt{\frac{\mu}{m}}(n+\tfrac{1}{2})$, where $n = 1, 3, 5, \ldots$. It can easily be seen that if the series used does not terminate then U becomes infinite as x increases, so the only admissible values of the energy are those given by $E_n = \frac{h}{2\pi}\sqrt{\frac{\mu}{m}}(n + \tfrac{1}{2})$, where $n = 0, 1, 2, 3, \ldots$. The classical frequency of vibration is equal to $\frac{1}{2\pi}\sqrt{\frac{\mu}{m}}$, so denoting this by ν_0 we have $E_n = h\nu_0(n + \tfrac{1}{2})$. The possible energies of the oscillator are therefore $\tfrac{1}{2}h\nu_0$, $\tfrac{3}{2}h\nu_0$, $\tfrac{5}{2}h\nu_0, \ldots$. According to this the energy cannot be zero. This result is different from that given by the older quantum theory, which was $E_n = nh\nu_0$.

With $E = \tfrac{1}{2}h\nu_0$ we get $U = a_0$, so that $w_0 = a_0e^{-ax^2}$. The chance of finding the particle between x and $x + dx$ is proportional to $\bar{w}w\,dx = a_0^2e^{-2ax^2}dx$. To make this equal to the chance we take $\int_{-\infty}^{+\infty} a_0^2 e^{-2ax^2}dx = 1$, which gives $a_0^2 = \sqrt{2a/\pi}$, so that the normalized value of w is $w = (2a/\pi)^{1/4}e^{-ax^2}e^{-\pi i\nu_0 t}$, where $a = \pi\sqrt{\mu m}/h$. With $E = \tfrac{3}{2}h\nu_0$ we get $U = a_1x$, so that $w_0 = a_1xe^{-ax^2}$.

With $E = (n + \tfrac{1}{2})h\nu_0$ the potential energy is equal to the total energy when $\tfrac{1}{2}\mu x^2 = (n + \tfrac{1}{2})h\nu_0$ or when $x = \sqrt{2(n + \tfrac{1}{2})h\nu_0/\mu}$. According to classical theory, this would be the maximum possible distance

from the origin, so that the chance of finding the particle at greater values of x would be zero. According to the present theory, the chance is proportional to $U^2e^{-2ax^2}$ and so does not become zero exactly even when x is very large. There is a finite chance of finding the particle at large values of x, where its potential energy is greater than its total energy, so that its kinetic energy would be negative, which is absurd. This result is interpreted as meaning that it is impossible to locate the particle in such positions without giving enough energy to make its kinetic energy positive.

20. Central Forces.

Schrödinger's equation $\Delta w + \frac{8\pi^2 m_0}{h^2}(E - V)w = 0$ in polar coordinates r, θ, ϕ is

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial w}{\partial r}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 w}{\partial\phi^2} + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial w}{\partial\theta}\right) + \frac{8\pi^2 m_0}{h^2}(E - V)w = 0.$$

If V is a function of r only, we can find a solution of the form $w = U(r)S(\theta, \phi)$.

Substituting this value of w, we find

$$\frac{1}{U}\frac{d}{dr}\left(r^2\frac{dU}{dr}\right) + \left[\frac{1}{S\sin^2\theta}\frac{\partial^2 S}{\partial\phi^2} + \frac{1}{S\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial S}{\partial\theta}\right)\right] + \frac{8\pi^2 m_0 r^2}{h^2}(E - V) = 0$$

The large bracket does not involve r and so must be a constant, say $-A$, so that

$$\frac{1}{S\sin^2\theta}\frac{\partial^2 S}{\partial\phi^2} + \frac{1}{S\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial S}{\partial\theta}\right) + A = 0.$$

Now let $S(\theta, \phi) = K(\phi)P(\theta)$, which gives

$$\frac{1}{K}\frac{d^2K}{d\phi^2} + \frac{\sin\theta}{P}\frac{d}{d\theta}\left(\sin\theta\frac{dP}{d\theta}\right) + A\sin^2\theta = 0.$$

The first term is independent of θ and so must be a constant. Let it equal $-m^2$, so that

$$\frac{d^2K}{d\phi^2} + m^2K = 0$$

and

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dP}{d\theta}\right) + \left(A - \frac{m^2}{\sin^2\theta}\right)P = 0.$$

When $\theta = 0$, $\dfrac{1}{\sin\theta}$ becomes infinite and there are only certain values of A for which P remains finite. It is well known that these are given by $A = l(l+1)$, where $l = 0, 1, 2, 3, \ldots$.

The equation $\dfrac{d^2K}{d\phi^2} + m^2K = 0$ gives $K = Ce^{im\phi}$, so that to make K a single valued function of ϕ we must have m an integer. With $C = 1/\sqrt{2\pi}$, K is normalized.

The functions P are the associated Legendre polynomials $P_{lm}(\theta)$ with $|m| \leq l$, of which the first few when normalized are

$$
\begin{aligned}
P_{00} &= 1/\sqrt{2} & P_{20} &= (\tfrac{5}{8})^{1/2}(3\cos^2\theta - 1) \\
P_{10} &= (\tfrac{3}{2})^{1/2}\cos\theta & P_{21} &= \tfrac{1}{2}(15)^{1/2}\sin\theta\cos\theta \\
P_{11} &= \frac{3^{1/2}}{2}\sin\theta & P_{22} &= \tfrac{1}{4}(15)^{1/2}\sin^2\theta
\end{aligned}
$$

$$
\begin{aligned}
P_{30} &= (\tfrac{7}{8})^{1/2}(5\cos^3\theta - 3\cos\theta) \\
P_{31} &= (\tfrac{21}{32})^{1/2}\sin\theta(5\cos^2\theta - 1) \\
P_{32} &= \tfrac{1}{4}(105)^{1/2}\sin^2\theta\cos\theta \\
P_{33} &= \tfrac{1}{4}(\tfrac{35}{2})^{1/2}\sin^3\theta.
\end{aligned}
$$

The radial function $U(r)$ is given by

$$\frac{1}{U}\frac{d}{dr}\left(r^2\frac{dU}{dr}\right) - A + \frac{8\pi^2 m_0 r^2}{h^2}(E - V) = 0$$

with $A = l(l+1)$. Putting $U(r) = u(r)/r$, we get

$$\frac{d^2u}{dr^2} - \frac{l(l+1)u}{r^2} + \frac{8\pi^2 m_0}{h^2}(E - V)u = 0.$$

21. Atoms with One Electron.

In the case of a nucleus with charge Ze with one electron, we have $V = -Ze^2/r$. Let $E = -2\pi^2 m_0 e^4 Z^2/n^2h^2$ and $r = nh^2x/8\pi^2 m_0 e^2 Z$ and substitute these in the equation for u, when it becomes

$$\frac{d^2u}{dx^2} + \left(-\frac{1}{4} + \frac{n}{x} - \frac{l(l+1)}{x^2}\right)u = 0.$$

Now let $u = sx^{l+1}e^{-x/2}$, which gives

$$x\frac{d^2s}{dx^2} + \{2(l+1) - x\}\frac{ds}{dx} + \{n - l - 1\}s = 0.$$

If $s = a_0 + a_1x + a_2x^2 + \ldots, a_jx^j + \ldots$, we find

$$a_{j+1}(j+1)(j+2l+2) = a_j(j+l+1-n).$$

This shows that the series ends and becomes a polynomial when $j + l + 1 = n$, and also that terms with negative powers of x cannot occur because with $j = -1$ we get $a_{-1} = 0$. Since $j = 0, 1, 2, 3, \ldots$, and $l = 0, 1, 2, 3, \ldots$, we see that n is a positive integer. If $l = 0$ then $n = 1, 2, 3, \ldots$, and if $l = 1$ then $n = 2, 3, 4, \ldots$, and if $l = 2$ then $n = 3, 4, 5, \ldots$, and so on. Thus it appears E can have the values $-2\pi^2 m_0 e^4 Z^2/n^2h^2$ where $n = 1, 2, 3, \ldots$. These are the same values of E as were given by Bohr's old quantum theory.

If we take $E = +2\pi^2 m_0 e^4 Z^2/n^2h^2$, then we get

$$\frac{d^2u}{dx^2} + \left(\frac{1}{4} + \frac{n}{x} - \frac{l(l+1)}{x^2}\right)u = 0.$$

This equation has finite solutions for any value of n, so that any positive value of E is possible. The solutions with positive E correspond to the hyperbolic orbits of the classical theory.

Some of the values of $U_{nl}(r)$ are as follows:

$$U_{10} = -2\left(\frac{Z}{r_0}\right)^{3/2} e^{-x/2} \qquad U_{20} = \frac{2}{2^{5/2}}\left(\frac{Z}{r_0}\right)^{3/2} e^{-x/2}(x-2)$$

$$U_{21} = -\frac{2}{2^2\sqrt{6}}\left(\frac{Z}{r_0}\right)^{3/2} e^{-x/2}x.$$

Here $r_0 = h^2/4\pi^2 m_0 e^2$ is equal to the radius of the orbit in a hydrogen atom on Bohr's theory with $n = 1$.

The integer $n = 1, 2, 3, \ldots$ which determines the energy is called the principal quantum number. The integer $l = 0, 1, 2, \ldots$ is called the azimuthal quantum number, and $m = 0, \pm 1, \pm 2, \ldots$ is sometimes called the magnetic quantum number. Its positive values are equal to or less than l. The following table gives the possible values of l and m with several values of n.

n	l	m	Total States
1	0	0	$2K$
2	0	0	
	1	1, 0, −1	$8L$
3	0	0	
	1	1, 0, −1	
	2	2, 1, 0, −1, −2	$18M$
4	0	0	
	1	1, 0, −1	
	2	2, 1, 0, −1, −2	
	3	3, 2, 1, 0, −1, −2, −3	$32N$

For each possible set of values of n, l and m there is a different "proper" function or "proper" state.* Thus with $n = 3$ there are 9

* German, *Eigenfunktion, Eigenzustand.*

different sets of values. Also an electron has two different proper functions for each set of values because its spin component can have two equal and opposite values. Therefore with $n = 3$, for example, there are 18 different possible proper states of the electron. According to Pauli's exclusion principle, only one electron can be in each proper state, so that in a many-electron atom the number of electrons with the quantum number n cannot be greater than twice the number of sets of values of n, l and m. The states with higher negative energy values are always filled up first in an unexcited atom. The states with $n = 1, 2, 3, 4, \ldots$ are called K, L, M, N states or energy levels respectively. In a many-electron atom the proper functions will differ more or less from the proper functions with only one electron, but it is supposed that the number of different proper functions is the same.

The grouping of the electrons in the inert gas atoms is supposed to be as follows; cf. Periodic Table, p. 433:

Gas	Z	K	L	M	N	O	P
Helium	2	2					
Neon	10	2	8				
Argon	18	2	8	8			
Krypton	36	2	8	18	8		
Xenon	54	2	8	18	18	8	
Radon	86	2	8	18	32	18	8

The alkali metals have one more electron than the inert gases, and the additional electron is supposed to be in a group by itself outside the other groups. Sodium has two K electrons and eight L electrons like neon, but has also one M electron.

In the same way the alkaline earth metals are supposed to have their electrons grouped like those of the inert gases, but with two additional electrons in an outer group.

The elements fluorine, chlorine, bromine and iodine have their electrons grouped like the inert gases, except that one electron is missing from the outermost group. The elements oxygen, sulphur, selenium and tellurium have two electrons missing as compared with the inert gases; and so on for the other elements.

The three cases considered are sufficient to show how Schrödinger's equation enables the possible energies of a particle moving in a field of force to be calculated. The solutions also give the chance of finding the particle in any element of volume dv, since this chance is $\bar{w}w\,dv$ when w has been normalized.

22. Operators.

An operator represents a rule which changes a function into another function. For example, let the operator δ represent differen-

tiation with respect to x. Then $\delta f(x) = \frac{df}{dx}$ and $\delta f(x, y) = \frac{\partial f}{\partial x}$. The vector operator $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$, so that $\Delta f(x, y, z) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$ is an important operator.

The sum of two operators α and β is defined by

$$(\alpha + \beta)f(x) = \alpha f(x) + \beta f(x),$$

and the product by

$$\alpha\beta f(x) = \alpha\big(\beta f(x)\big).$$

The product $\alpha\beta$ may not be equal to the product $\beta\alpha$. For example let

$$\alpha f(x) = xf(x) \quad \text{and} \quad \beta f(x) = \big(f(x)\big)^2.$$

Then
$$\alpha\beta f(x) = \alpha\big(f(x)\big)^2 = x\big(f(x)\big)^2,$$

but
$$\beta\alpha f(x) = \beta\big(xf(x)\big) = x^2\big(f(x)\big)^2.$$

If $\alpha\beta f(x) = \beta\alpha f(x)$, or $\alpha\beta = \beta\alpha$, then the operators α and β are said to commute.

If a function $f(x)$ is such that an operator α gives $\alpha f(x) = af(x)$, where a is a constant, then $f(x)$ is called a characteristic or proper function of the operator α and a is called a characteristic or proper value of α belonging to the proper function $f(x)$.

For example, $\delta e^{kx} = \frac{\partial}{\partial x} e^{kx} = ke^{kx}$, so that e^{kx} is a proper function of the operator δ and k is a proper value of δ.

In quantum mechanics the functions of the co-ordinates are required to be single-valued and continuous, and to give a finite result when the square of the absolute value is integrated over the whole range of the variables. Thus, a function $\phi(x, y, z)$ must give

$$\int \bar{\phi}\phi \, dx\,dy\,dz = c$$

where c is a finite constant. ϕ may be infinite at a finite number of points but $\int \bar{\phi}\phi \, dx\,dy\,dz$ must be finite. Such functions will be said to belong to class Q. $\bar{\phi}$ is the conjugate complex of ϕ so that if $\phi = a + ib$, then $\bar{\phi} = a - ib$.

Consider two functions, ϕ_1 and ϕ_2, belonging to class Q, and let an operator α give $\alpha\phi_2 = a\phi_2$, so that a is a proper value of α belonging to the proper function ϕ_2.

The operator α is said to be Hermitian if $\int \bar{\phi}_1 \alpha \phi_2 dv = \int \phi_2 \bar{\alpha} \bar{\phi}_1 dv$. The integrals are supposed to cover the whole range of the variables, and dv is an element of volume in the co-ordinate space. For example, if ϕ_1 and ϕ_2 are functions of x, y, z, then $dv = dx dy dz$.

If α is Hermitian, then its proper values are real numbers. This may be shown as follows:

The equation $\alpha\phi_2 = a\phi_2$ gives $\bar{\alpha}\bar{\phi}_2 = \bar{a}\bar{\phi}_2$, so that

$$\int \bar{\phi}_2 \alpha \phi_2 dv = a \int \bar{\phi}_2 \phi_2 dv,$$

and

$$\int \phi_2 \bar{\alpha} \bar{\phi}_2 = \bar{a} \int \phi_2 \bar{\phi}_2 dv.$$

But if α is Hermitian, the two integrals on the left are equal so that $a = \bar{a}$, and therefore a is real.

An operator α is said to be linear if

$$\alpha\{c_1 f_1(x) + c_2 f_2(x)\} = c_1 \alpha f_1(x) + c_2 \alpha f_2(x),$$

where c_1 and c_2 are constants.

23. General Principles of Quantum Mechanics.

In classical dynamics the state of a system with f degrees of freedom, such as an atom, is determined by the values, at a time t, of the f co-ordinates $q_1, q_2, \ldots q_f$ and the f conjugate momenta $p_1, p_2 \ldots p_f$.

In quantum mechanics no such precise specification of the state of the system is possible. The state of the system is supposed determined as closely as possible by a state function w of the co-ordinates $q_1, q_2, \ldots q_f$ and the time t, or $w(q_1, q_2 \ldots q_f, t)$.

The chance of finding q_1 between q_1 and $q_1 + dq_1$, q_2 between q_2 and $q_2 + dq_2$, and so on, is proportional to $\bar{w}w\,dv$, where

$$dv = dq_1 dq_2 \ldots dq_f.$$

If $\int \bar{w}w\,dv = 1$, then the chance is equal to $\bar{w}w\,dv$, and w is said to be normalized.

The state function is a function of the co-ordinates and the time, but its value is not determined by the positions of the particles which are unknown. A point P is selected with co-ordinates $q_1, q_2, q_3, \ldots q_f$, in the co-ordinate space and dv is an element of volume at P. Then $\bar{w}w\,dv$, when the values of the co-ordinates of P and the time t are used in $\bar{w}$ and w, gives the chance of finding the particles in dv at the time t. The partial differential coefficient of w with respect to t or $\partial w/\partial t$ is the rate of variation of w at a fixed point P. If the point P

were supposed to be moving with velocity components $\dot{q}_1, \dot{q}_2, \ldots \dot{q}_f$, then we should have

$$\frac{dw}{dt} = \frac{\partial w}{\partial q_1}\dot{q}_1 + \frac{\partial w}{\partial q_2}\dot{q}_2 + \ldots + \frac{\partial w}{\partial t},$$

but in this expression $\dot{q}_1, \dot{q}_2$, would not be the velocities of the particles, and would have no relation to them. Usually the point P is supposed fixed, and no useful purpose is served by supposing it to be moving.

Suppose we have a free particle with momentum p_x and energy E moving along the x-axis. The state function w for this particle must represent waves of frequency ν given by $E = h\nu$ and wavelength λ given by $\lambda = h/p_x$. We may take $w = Ae^{i(2\pi/\lambda)(x - ut)}$, where u is the wave velocity. Putting $1/\lambda = p_x/h$ and $u/\lambda = E/h$, this gives $w = Ae^{(2\pi i/h)(p_x x - Et)}$. Differentiating with respect to x, we get $\partial w/\partial x = 2\pi i p_x w/h$, so that $p_x w = \frac{h}{2\pi i}\frac{\partial w}{\partial x}$. Thus we may regard p_x as an operator which, operating on w, is equivalent to $\frac{h}{2\pi i}\frac{\partial}{\partial x}$. In the same way, $p_y = \frac{h}{2\pi i}\frac{\partial}{\partial y}$ and $p_z = \frac{h}{2\pi i}\frac{\partial}{\partial z}$. It is assumed that any function of the co-ordinates and momenta may be converted into an operator operating on a state function w by replacing p_x, p_y and p_z by $\frac{h}{2\pi i}\frac{\partial}{\partial x}$, $\frac{h}{2\pi i}\frac{\partial}{\partial y}$ and $\frac{h}{2\pi i}\frac{\partial}{\partial z}$, respectively, or by replacing p_n by $\frac{h}{2\pi i}\frac{\partial}{\partial q_n}$.

According to this, any dynamical variable of a dynamical system which is equal to a function of the co-ordinates and momenta can be represented by a linear operator. If the operator is Hermitian, the proper values of the variable will be real. Since only real values of observable quantities need be considered, the linear operators should be Hermitian. The order of the factors in the function representing the variable should, therefore, if necessary, be arranged so as to make the operator Hermitian.

The most important dynamical variable is the energy. The energy operator is obtained from the Hamiltonian expression for the energy as a function of the co-ordinates and momenta or $H(q, p, t)$, where q and p stand for all the q's and all the p's. The energy operator is then $H\left(q, \frac{h}{2\pi i}\frac{\partial}{\partial q}, t\right)$.

The Hamiltonian expression for the energy H is equal to E, so that when H is regarded as an operator, we have

$$H\left(q, \frac{h}{2\pi i}\frac{\partial}{\partial q}, t\right)w = Ew.$$

The state function w, of course, is a function of the co-ordinates q and the time t or $w(q, t)$.

In the case of a single particle, of mass m, moving in a field of force with potential V, we have

$$H = \frac{p_x^2 + p_y^2 + p_z^2}{2m} + V = E.$$

This gives, putting $\quad p_x^2 = \left(\frac{h}{2\pi i}\frac{\partial}{\partial x}\right)^2$, &c.,

$$H = \frac{1}{2m}\left(\frac{h}{2\pi i}\right)^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + V = E,$$

so that $Hw = Ew$ becomes

$$-\frac{1}{2m}\frac{h^2}{4\pi^2}\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}\right) = (E - V)w,$$

or
$$\Delta w + \frac{8\pi^2 m}{h^2}(E - V)w = 0,$$

which is Schrödinger's equation.

If the equation for a free particle $w = Ae^{(2\pi i/h)(p_x x - Et)}$ is differentiated partially with respect to t, we get $\frac{\partial w}{\partial t} = -\frac{2\pi i}{h}Ew$, or $Ew = -\frac{h}{2\pi i}\frac{\partial w}{\partial t}$. This suggests that the energy operator H may be equivalent to $-\frac{h}{2\pi i}\frac{\partial}{\partial t}$ just as $p_n = \frac{h}{2\pi i}\frac{\partial}{\partial q_n}$. It is assumed, therefore, in quantum mechanics that $Hw = -\frac{h}{2\pi i}\frac{\partial w}{\partial t}$, where $w = w(q, t)$ may be any function of the co-ordinates and the time.

With $\quad H = \frac{p_x^2 + p_y^2 + p_z^2}{2m} + V, \quad Hw = -\frac{h}{2\pi i}\frac{\partial w}{\partial t}$

gives
$$\Delta w - \frac{8\pi^2 m}{h^2}Vw + \frac{4\pi m i}{h}\frac{\partial w}{\partial t} = 0.$$

This form of Schrödinger's equation holds for any w. The equation $Hw = Ew$ only holds for w's that are proper functions for H belonging to the proper values of the energy E.

When $Hw = Ew$, then $Hw = -\frac{h}{2\pi i}\frac{\partial w}{\partial t}$ gives $Ew = -\frac{h}{2\pi i}\frac{\partial w}{\partial t}$, which gives $w(q, t) = w(q)e^{-2\pi i Et/h}$, where $w(q, t)$ is one of the proper functions for H.

For the proper states with $w = w(q)e^{-2\pi iEt/h}$,

$$\bar{w}w\,dv = \bar{w}(q)e^{2\pi iEt/h}\,w(q)e^{-2\pi iEt/h}\,dv = \bar{w}(q)w(q)\,dv.$$

The chance of finding the particles in dv is therefore constant, so the proper states of the energy operator are called stationary states.

When an atom is in a stationary state, $Hw = Ew$ and an observation of the energy E will certainly give the value E, provided the observation is made with no unnecessary disturbance of the system.

If the state function w is a proper function of an operator α representing a dynamical variable A, then $\alpha w = aw$, where a is the proper value of A belonging to the proper function w. When A is the energy so that $\alpha = H$, then $Hw = Ew$.

The operator for the momentum p_n does not commute with the conjugate co-ordinate q_n. Thus

$$p_n q_n w = \frac{h}{2\pi i}\frac{\partial}{\partial q_n}(q_n w) = \frac{h}{2\pi i}\left(w + q_n\frac{\partial w}{\partial q_n}\right)$$

and

$$q_n p_n w = q_n \frac{h}{2\pi i}\frac{\partial w}{\partial q_n},$$

so that

$$(p_n q_n - q_n p_n)w = \frac{h}{2\pi i}\,w.$$

Since this holds for any w, we may write

$$p_n q_n - q_n p_n = \frac{h}{2\pi i}.$$

Here, of course, it is understood that $p_n q_n - q_n p_n$ is to be regarded as an operator with $p_n = \frac{h}{2\pi i}\frac{\partial}{\partial q_n}$. Note that if p and q are not conjugate, then $\frac{\partial}{\partial q_n}(q_m w) = q_m\frac{\partial w}{\partial q_n}$ because $\frac{\partial q_m}{\partial q_n} = 0$, so that

$$p_n q_m - q_m p_n = 0,\ n \neq m.$$

24. **Orthogonal Functions.**

Two functions, $f_1(x)$ and $f_2(x)$, are said to be orthogonal in the interval from $x = a$ to $x = b$, if

$$\int_a^b \overline{f_1(x)} f_2(x)\,dx = 0.$$

A set of functions $f_1, f_2, f_3, f_4, \ldots$, is an orthogonal set if $\int_a^b \bar{f}_n f_m dx = 0$, $n \neq m$. If $\int_a^b \bar{f}_n f_n dx = 1$ for all values of n the set is normalized.

Any function F of class Q can be expanded in a series of orthogonal functions, thus

$$F(x) = c_1 f_1(x) + c_2 f_2(x) + \ldots = \sum_n c_n f_n(x).$$

To determine the constants c_n multiply by $\overline{f_m(x)}$ and integrate from a to b.

$$\int_a^b F(x)\overline{f_m(x)}\, dx = \sum_n c_n \int_a^b f_n(x)\overline{f_m(x)}\, dx.$$

But $\int_a^b f_n(x)\overline{f_m(x)}\, dx = 0$, unless $n = m$, so that

$$\int_a^b F(x)\overline{f_m(x)}\, dx = c_m \int_a^b \overline{f_m(x)} f_m(x)\, dx.$$

If $f_m(x)$ is normalized,

$$\int_a^b \overline{f_m(x)} f_m(x)\, dx = 1,$$

so that
$$c_m = \int_a^b F(x)\overline{f_m(x)}\, dx.$$

Now consider the proper functions w_n of an Hermitian operator α. We have $\alpha w_n = a_n w_n$ where $a_1, a_2, a_3, \ldots$ are the proper values of the variable A represented by α. We have

$$\int \bar{w}_n \alpha w_m dv = a_m \int \bar{w}_n w_m dv,$$

and
$$\int w_m \bar{\alpha} \bar{w}_n dv = \bar{a}_n \int w_m \bar{w}_n dv,$$

since $\bar{\alpha}\bar{w}_n = \bar{a}_n \bar{w}_n$. But $\int \bar{w}_n \alpha w_m dv = \int w_m \bar{\alpha} \bar{w}_n dv$, when α is Hermitian and $\bar{a}_n = a_n$, so that

$$a_m \int \bar{w}_n w_m dv = a_n \int \bar{w}_n w_m dv,$$

or
$$(a_n - a_m) \int \bar{w}_n w_m dv = 0.$$

Therefore, if $a_n \neq a_m$, we get $\int \bar{w}_n w_m dv = 0$, so that w_n and w_m are orthogonal functions. The proper functions of any Hermitian operator, therefore, form an orthogonal set.

Any state function w can be expanded in a series of the proper functions of an operator α Thus

$$w = c_1 w_1 + c_2 w_2 + c_3 w_3 + \ldots,$$

because the proper functions $w_1, w_2, \ldots$ are orthogonal. The coefficients are given by $c_n = \int w \bar{w}_n dv$ provided the w_n are normalized.

If α is the energy operator H, then if $Hw_n = E_n w_n$, the E_n are the proper values of the energy. If the energy is determined by an observation, then if $w = w_n$, the energy will certainly be found to have the value E_n. When w is not one of the proper functions of H, then it is supposed that an observation of the energy will not always give the same result, but will always give one of the proper values E_n. The chance of getting any particular proper energy E_n will depend on w. If w is only slightly different from, say, w_2, then the chance of getting E_2 will be nearly equal to unity.

If an observation of the energy gives E_n, then an immediate repetition of the observation will certainly give the same result, because there will not have been time for the state to change. Therefore, the observation of the energy changes the state from w to w_n, so that after an observation giving E_n, the atom will be in the state w_n, whatever the state function w was before the observation.

We have $w = \sum\limits_n c_n w_n$, so that

$$\bar{w} = \sum_n \bar{c}_n \bar{w}_n,$$

and

$$\int \bar{w} w \, dv = \int \sum_n c_n \bar{w}_n \sum_n c_n w_n dv = 1,$$

if w is normalized. This gives

$$\bar{c}_1 c_1 + \bar{c}_2 c_2 + c_3 c_3 + \ldots = 1,$$

if w_n is normalized, because then

$$\int \bar{w}_n w_m dv = 0, \quad n \neq m,$$

and

$$\int \bar{w}_n w_n dv = 1.$$

It is assumed that when the atom is in a state with normalized state function $w = \sum\limits_n c_n w_n$, then the chance of an observation giving $E = E_n$ is equal to $\bar{c}_n c_n$, with w_n states also normalized.

One of the possible energy values E_n must be found so that the sum of the chances must be unity and, as we have just seen, $\sum\limits_n \bar{c}_n c_n = 1$.

It follows that if a large number of observations are made on

atoms all in the same state w, then the average value $\bar{E}$ of the energies found will be given by

$$\bar{E} = \sum_n \bar{c}_n c_n E_n.$$

In classical theory an atom in a given state has a definite energy. In quantum mechanics this is not the case. An atom with a state function $w = \sum_n c_n w_n$ is considered to be in a definite state determined by w, but it cannot be said to have a definite energy. But the average of a large number of observations on atoms all in the state w has a definite value. It is important to remember that an observation giving E_n changes w to w_n. The large number of observations are all supposed to be made on atoms in state w and so cannot be made on the same atom unless it is supposed that after each observation the state of the atom is changed back from w_n to the original w.

We have supposed that w was expanded in a series of proper functions of the energy operator H, but any other operator could have been used in the same way. Thus, if w is expanded in a series of proper functions of an operator α representing any dynamical variable A, then $w = \sum_n c_n w_n$ and an observation of the value of A will give one of the proper values of A. The chance of getting A_n will be $\bar{c}_n c_n$, and the average value for a large number of observations will be $\sum_n \bar{c}_n c_n A_n$, just as with the energy operator.

Now consider the integral $\int \bar{w} \alpha w \, dv$ taken as usual over the whole range of the co-ordinates. Expanding w in a series of proper functions belonging to α, we have $w = \sum_n c_n w_n$, and $\bar{w} = \sum_n \bar{c}_n \bar{w}_n$, so that the integral becomes $\int \sum_n \bar{c}_n \bar{w}_n \sum c_n \alpha w_n \, dv$; but $\alpha w_n = A_n w_n$, where A_n is a proper value of A. The integral is therefore equal to

$$\int \sum_n \bar{c}_n \bar{w}_n \sum_n c_n A_n w_n \, dv,$$

which, since $\int \bar{w}_n w_m \, dv = 0$, $n \neq m$, and $\int \bar{w}_n w_n \, dv = 1$, gives $\sum_n A_n \bar{c}_n c_n$. But this, as we have seen, is just the average value of a large number of observations of A, all made on atoms in the state w. It appears, therefore, that $\int \bar{w} \alpha w \, dv$ is equal to the average value of the variable A for the state w.

25. Matrices.

Consider two operators α and β representing two variables A and B. Let the proper functions of α be u_n, and those of β be w_n, so that

$\alpha u_n = a_n u_n$, and $\beta w_n = b_n w_n$, where a_n and b_n are the proper values of A and B respectively.

Now suppose the state function of an atom is αw_n. Expanding this in a series of proper functions of β, we get $\alpha w_n = \sum_m c_{mn} w_m$. The coefficients c_{mn} depend on w_n and w_m, as indicated by the two suffixes. Multiplying by $\bar{w}_p$ and integrating gives

$$\int \bar{w}_p \alpha w_n dv = \sum_m c_{mn} \int \bar{w}_p w_m dv = c_{pn},$$

because

$$\int \bar{w}_p w_m dv = 0,\ p \neq m,\ \text{and} = 1,\ p = m.$$

The values of $c_{nm} = \int \bar{w}_n \alpha w_m dv$ may be arranged in rows and columns thus:

$$\begin{matrix} c_{11} & c_{12} & c_{13} & c_{14} \dots \\ c_{21} & c_{22} & c_{23} & c_{24} \dots \\ c_{31} & c_{32} & c_{33} & c_{34} \dots \\ \dots & & \dots & \dots \end{matrix}$$

Such a set of numbers is called a *matrix*, and the above matrix is called the matrix of the operator α with the proper functions of the operator β.

If α is an Hermitian operator,

$$c_{nm} = \int \bar{w}_n \alpha w_m dv = \int w_m \bar{\alpha} \bar{w}_n dv = \bar{c}_{mn}.$$

A matrix for which $c_{nm} = \bar{c}_{mn}$ is called an Hermitian matrix, so the matrices of Hermitian operators are Hermitian matrices.

It is convenient to use the same symbol for the terms of the matrix as for the operator. Thus $\alpha_{nm} = \int \bar{w}_n \alpha w_m dv$ is the nm term of the matrix for the operator α. α_{nm}, of course, is just a number.

If $\alpha = \beta$ as a particular case, then

$$\alpha_{nm} = \int \bar{w}_n \beta w_m dv = \int \bar{w}_n b_m w_m dv = b_m \int \bar{w}_n w_m dv,$$

so that

$$\alpha_{nm} = 0,\ n \neq m,\ \text{and}\ \ \alpha_{nm} = b_m,\ n = m.$$

The matrix in this case is

$$\begin{matrix} b_1 & 0 & 0 & 0 \dots \\ 0 & b_2 & 0 & 0 \dots \\ 0 & 0 & b_3 & 0 \dots \\ \dots & & \dots & \dots \end{matrix}$$

Such a matrix is called a diagonal matrix because all the terms are zero except the diagonal ones.

The most important matrices are those of an operator α with the proper functions of the energy operator H. With α also equal to H,

$$\alpha_{nm} = \int \bar{w}_n H w_m dv = E_n \int \bar{w}_n w_m dv,$$

so that $\alpha_{nm} = E_m$, $n = m$, and $\alpha_{nm} = 0$, $n \neq m$. The energy matrix is then

$$\begin{matrix} E_1 & 0 & 0 & 0 & 0 \dots \\ 0 & E_2 & 0 & 0 & 0 \dots \\ 0 & 0 & E_3 & 0 & 0 \dots \\ \cdot & \cdots & \cdots & & \cdots \end{matrix}$$

so that the diagonal terms are equal to the proper values of the energy.

In the matrix for an operator α with the proper functions w_n of the energy operator H, or $\alpha_{nm} = \int \bar{w}_n \alpha w_m dv$, we may put

$$\bar{w}_n = \bar{w}_{n0} e^{2\pi i E_n t/h} \text{ and } w_m = w_{m0} e^{-2\pi i E_m t/h},$$

where $\bar{w}_{n0}$ and w_{m0} are functions of the co-ordinates only and independent of the time.

This gives

$$\alpha_{nm} = e^{-(2\pi i/h)(E_m - E_n)t} \int \bar{w}_{n0} \alpha w_{m0} dv.$$

If α is Hermitian, then $\bar{\alpha}_{nm} = \alpha_{mn}$, so that

$$\alpha_{nm} + \alpha_{mn} = \alpha_{nm} + \bar{\alpha}_{nm}$$

is real.

With $$\alpha_{nm} = e^{-(2\pi i/h)(E_m - E_n)t} \int \bar{w}_{n0} \alpha w_{m0} dv,$$

this gives

$$\alpha_{nm} + \alpha_{mn} = 2 \left| \int \bar{w}_{n0} \alpha w_m dv \right| \cos\left\{\frac{2\pi}{h}(E_m - E_n)t + \theta\right\},$$

which represents a real oscillation in the atom with frequency $\nu = (E_m - E_n)/h$ and amplitude $2 \left| \int \bar{w}_{n0} \alpha w_{m0} dv \right|$

If $n = m$, then $\nu = 0$, so that the diagonal terms α_{nm} of the matrix are constants

In Bohr's quantum theory it was supposed that when an atom with energy E_m makes a transition to a state with energy E_n, the energy difference $E_m - E_n$ is emitted as a photon with frequency ν given by $\nu = (E_m - E_n)/h$. This assumption is also made in quantum mechanics. It appears that the frequencies of the matrix elements

α_{nm} are the same as the frequencies of the photons emitted by the atom

As a particular case, let α represent the x component of the electric moment of the atom. Then α_{nm} represents an oscillating electric moment of frequency $(E_m - E_n)/h$ and amplitude

$$2\left|\int \bar{w}_{n0}\alpha w_{m0}dv\right| = 2\alpha_{nm0}, \text{ say.}$$

According to classical electromagnetic theory, such an oscillating electric moment radiates energy equal to $\frac{4}{3}\frac{(2\pi\nu)^4}{C^3}\,|\,\alpha_{nm0}\,|^2$ ergs per second. The number of photons with energy $h\nu$ emitted per second is therefore $\frac{4}{3}\frac{(2\pi\nu)^4}{C^3 h\nu}\,|\,\alpha_{nm0}\,|^2$. The chance that a photon is emitted in a time dt is therefore $\frac{1}{3}\frac{(2\pi\nu)^4}{C^3 h\nu}\,|\,\alpha_{nm0}\,|^2 dt$.

Heisenberg supposed that this would be the chance that an atom with energy E_m would change to a state with energy E_n, emitting a photon with energy $h\nu$ in the time interval dt.

Thus, if $\alpha_{nm0} = 0$, there will be no radiation of frequency $\nu = (E_m - E_n)/h$. This result is important because α_{nm0} is often zero, so that many transitions which might be expected to occur do not, in fact, take place.

The matrix of the product of two operators α and β can be found as follows. Let $\alpha w_n = \sum\limits_m \alpha_{mn} w_m$, where $\alpha_{mn} = \int \bar{w}_m \alpha w_n dv$, and $\beta w_n = \sum\limits_m \beta_{mn} w_m$, where $\beta_{mn} = \int \bar{w}_m \beta w_n dv$. Also, let $\gamma = \alpha\beta$, and $\gamma w_n = \sum\limits_m \gamma_{mn} w_m$, where $\gamma_{mn} = \int \bar{w}_m \gamma w_n dv$.

Then $\gamma w_n = \alpha\beta w_n = \alpha\sum\limits_k \beta_{kn} w_k = \sum\limits_k \beta_{kn}\alpha w_k$. But $\alpha w_k = \sum\limits_m \alpha_{mk} w_m$, so

$$\gamma w_n = \sum_k \beta_{kn} \sum_m \alpha_{mk} w_m = \sum_m\sum_k \alpha_{mk}\beta_{km} w_m,$$

so that

$$\sum_m \gamma_{mn} w_m = \sum_m\sum_k \alpha_{mk}\beta_{kn} w_m,$$

and therefore

$$\gamma_{mn} = \sum_k \alpha_{mk}\beta_{kn} = (\alpha\beta)_{mn}.$$

In the same way $(\beta\alpha)_{mn} = \sum\limits_k \beta_{mk}\alpha_{kn}$, so that $(\alpha\beta)_{mn}$ is not equal to $(\beta\alpha)_{mn}$, unless $\alpha\beta = \beta\alpha$.

It sometimes happens that two operators have the same proper functions. Let α and β be the operators and let $\alpha w_n = a_n w_n$, and $\beta w_n = b_n w_n$, where the w's are the same functions in both cases.

Then $$\alpha\beta w_n = \alpha b_n w_n = b_n \alpha w_n = b_n a_n w_n.$$
Also $$\beta\alpha w_n = \beta a_n w_n = a_n \beta w_n = a_n b_n w_n.$$

Now let ϕ be an arbitrary function, and let $\phi = \Sigma c_n w_n$.

Then $$\alpha\beta\phi = \Sigma c_n a_n b_n w_n,$$
and $$\beta\alpha\phi = \Sigma c_n a_n b_n w_n,$$
so that $$(\alpha\beta - \beta\alpha)\phi = 0,$$

and therefore $\alpha\beta = \beta\alpha$, so that α and β commute Thus, if two operators have the same set of proper functions, they commute

The converse theorem may also be proved; namely, that if two operators commute, then a set of orthogonal functions can be found which are proper functions for both operators.

For example, the proper functions of H, the energy operator, are also proper functions for the operator M representing the angular momentum. H and M therefore commute so that $HM - MH = 0$ In such an equation between operators, the operators may be replaced by the matrices representing them. The product of the matrices H_{nm} and M_{nm}, as we have seen, is given by $(HM)_{nm} = \sum\limits_k H_{nk} M_{km}$, so the equation $HM - MH = 0$ gives $\sum\limits_k H_{nk} M_{km} - \Sigma M_{nk} H_{km} = 0$; but $H_{nm} = 0$ unless $n = m$, so that we get

$$H_{nn} M_{nm} - M_{nm} H_{mm} = 0,$$
or $$M_{nm}(H_{nn} - H_{mm}) = 0;$$

therefore, $M_{nm} = 0$ unless $n = m$. The matrix for M is, therefore, a diagonal matrix. M is therefore a constant, because the diagonal terms of any matrix $\alpha_{nm} = \int \bar{w}_n \alpha w_m \, dv$, where the w's are proper functions of the energy operator H, are constants.

Thus we see that any dynamical variable whose operator commutes with H is a constant. Its possible values are those of the diagonal terms in its matrix.

26. The Variation Method.

Schrödinger's equation $Hw = Ew$ can be solved, as we have seen, in a few simple cases. That is, the proper functions w_n belonging to the proper values of E or E_n can be found.

In many cases it is not possible to find the proper functions exactly, but approximate values can be found. The variation method is a useful method often used to find approximate values of the lowest proper value of E, and the corresponding proper function.

Consider the integral $\int \phi H \phi dv$ over the whole range of the variables where H is the energy operator and ϕ any function of class Q, normalized so that $\int \phi\phi dv = 1$.

Let $\phi = \sum_n c_n w_n$, so that the integral is $\int \sum_m \bar{c}_m \bar{w}_m H \sum_n c_n w_n dv$. But $Hw_n = E_n w_n$, so the integral is $\int \sum_m \bar{c}_m \bar{w}_m \sum_n c_n w_n E_n dv$. But

$$\int \bar{w}_m w_n dv = 0, m \neq n, \text{ and } \int \bar{w}_n w_n dv = 1,$$

so

$$\int \phi H \phi dv = \sum_n \bar{c}_n c_n E_n.$$

Now $c_n c_n$ is positive, so that $\sum_n \bar{c}_n c_n E_n$ must be greater than the lowest value of E or E_1. We have, therefore,

$$\int \bar{\phi} H \phi dv > E_1.$$

If $\phi = w_1$, then $$\int \bar{\phi} H \phi dv = \int \bar{w}_1 H w_1 dv = E_1,$$

so that

$$\int \phi H \phi dv \geq E_1.$$

If ϕ is not normalized, then

$$\int \bar{\phi} H \phi dv \geq E_1 \int \bar{\phi}\phi dv.$$

If, therefore, a function ϕ is chosen so as to be some sort of approximation to w_1, and containing a variable parameter or parameters, and the parameters are varied so as to make $\int \phi H \phi dv$ as small as possible, then the minimum value of $\int \phi H \phi dv$ will be some sort of approximation to E_1 and the value of ϕ for the minimum value of the integral will be an approximation to w_1.

For example, ϕ may be taken to be given by $\phi = \sum_n c_n u_n$ where the u_n are functions believed to resemble w_1. The coefficients c_n are then varied so as to make $\int \phi H \phi dv$ a minimum and $\int \bar{\phi} H \phi dv$ will then be an approximation to E_1 with ϕ normalized.

Let

$$\int \bar{u}_n u_m dv = \Delta_{nm}, \quad \int \bar{u}_n H u_m dv = H_{nm}$$

and

$$E = \int \phi H \phi dv \Big/ \int \bar{\phi}\phi dv.$$

Then, putting $\phi = \sum_{n=1}^{n} c_n u_n$ in the last equation, we get

$$E\sum_{nm}\bar{c}_n c_m \Delta_{nm} = \sum_{nm}\bar{c}_n c_m H_{nm}$$

Differentiating this with respect to c_n and putting $\partial E/\partial c_n = 0$ gives

$$E\sum_{n}\bar{c}_n \Delta_{nk} = \sum_{n}\bar{c}_n H_{nk}$$

Also differentiating with respect to $\bar{c}_n$, and putting $\partial E/\partial \bar{c}_n = 0$, gives

$$E\sum_{n} c_n \Delta_{nk} = \sum_{n} c_n H_{nk}$$

Both equations are equivalent to

$$\sum_{n} c_n(H_{kn} - \Delta_{kn}E) = 0, \; k = 1, 2, 3 \; . \; . \; n.$$

Eliminating the c's, these equations give the determinant

$$\begin{vmatrix} H_{11} - \Delta_{11}E, & H_{12} - \Delta_{12}E, & \ldots & H_{1n} - \Delta_{1n}E \\ H_{21} - \Delta_{21}E, & H_{22} - \Delta_{22}E, & \ldots & \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ H_{n1} - \Delta_{n1}E_1, & \ldots & \ldots & H_{nn} - \Delta_{nn}E \end{vmatrix} = 0.$$

This is an equation of the nth degree in E with n roots The lowest root will be an approximation to E_1 because $\partial E/\partial c_n$ and $\partial E/\partial \bar{c}_n$ have been put equal to zero, so making $\int\bar{\phi}H\phi dv$ a minimum.

The variation method may be used to get an approximate solution of the problem of the hydrogen molecular ion H_2^+ in its state of lowest energy.

The H_2^+ ion consists of two protons at a distance d apart and one electron at a distance r_1 from one proton and r_2 from the other. The Hamiltonian energy operator is

$$H = -\frac{h^2}{8\pi^2 m}\Delta - \frac{e^2}{r_1} - \frac{e^2}{r_2} + \frac{e^2}{d}.$$

If the electron is very near one or other of the protons, H will be nearly the same as H for a hydrogen atom, so we may suppose

$$\phi = c_1 u_1 + c_2 u_2,$$

where u_1 and u_2 are the state functions for a hydrogen atom with its lowest energy, u_1 with the electron near one of the protons, and u_2 with it near the other one.

The determinantal equation is then

$$\begin{vmatrix} H_{11} - \Delta_{11}E, & H_{12} - \Delta_{12}E \\ H_{21} - \Delta_{21}E, & H_{22} - \Delta_{22}E \end{vmatrix} = 0,$$

which has two roots E_1 and E_2.

$\Delta_{11} = \Delta_{22} = 1$ if u_1 and u_2 are normalized. Also, $H_{12} = H_{21}$, so we get

$$E_1 = \frac{H_{11} + H_{12}}{1 + \Delta_{12}}, \text{ and } E_2 = \frac{H_{11} - H_{12}}{1 - \Delta_{12}}.$$

The two equations

$$c_1(H_{11} - \Delta_{11}E) + c_2(H_{12} - \Delta_{12}E) = 0$$

and

$$c_1(H_{21} - \Delta_{21}E) + c_2(H_{22} - \Delta_{22}E) = 0$$

with $E = E_1$, give $c_1 = c_2$, so that $\phi_1 = c_1(u_1 + u_2)$; and with $E = E_2$ give $c_2 = -c_1$, so that

$$\phi_2 = c_2(u_1 - u_2).$$

The integrals Δ_{12}, H_{11} and H_{12} may be calculated.

$$u_1 = e^{-r_1/a_0}/\sqrt{(\pi a_0^3)} \text{ and } u_2 = e^{-r_2/a_0}/\sqrt{(\pi a_0^3)},$$

where $a_0 = h^2/4\pi^2 me^2 = 0{\cdot}53 \times 10^{-8}$ cm. is the radius of the electron orbit nearest to the proton on Bohr's theory of the hydrogen atom.

It is found that

$$\Delta_{12} = e^{-\rho}\left(1 + \rho + \frac{\rho^2}{3}\right) \text{ where } \rho = d/a_0.$$

$$H_{11} = E_{\mathrm{H}} + \frac{e^2}{\rho a_0} + J,$$

where E_{H} is the lowest energy of the hydrogen atom and

$$J = -\int u_1 \frac{e^2}{r_2} u_1 dv = -\frac{e^2}{\rho a_0}\Big(1 - e^{-2\rho}(1 + \rho)\Big),$$

$$H_{12} = \left(E_{\mathrm{H}} + \frac{e^2}{\rho a_0}\right)\Delta_{12} + K,$$

where

$$K = -\int u_1 \frac{e^2}{r_1} u_2 dv = -\frac{e^2}{a_0} e^{-\rho}(1 + \rho).$$

These results give

$$E_1 = E_{\mathrm{H}} + \frac{e^2}{\rho a_0} + \frac{J + K}{1 + \Delta_{12}}$$

and
$$E_2 = E_H + \frac{e^2}{\rho a_0} + \frac{J - K}{1 - \Delta_{12}}.$$
K is negative, so E_1 is the lower of the two energy values.

When $\rho = d/a_0$ is large, $E_1 = E_H$ as would be expected, because, when d is large, the electron in its state of lowest energy must be near one of the protons forming a hydrogen atom. When ρ is very small, $\frac{J + K}{1 + \Delta_{12}}$ is equal to $-e^2/a_0$, or $2E_H$, so that $E_1 = 3E_H + \frac{e^2}{a_0 \rho}$. When ρ is very small, the two protons and the electron should have energy $4E_H + \frac{e^2}{a_0 \rho}$, so that with ρ small the variation

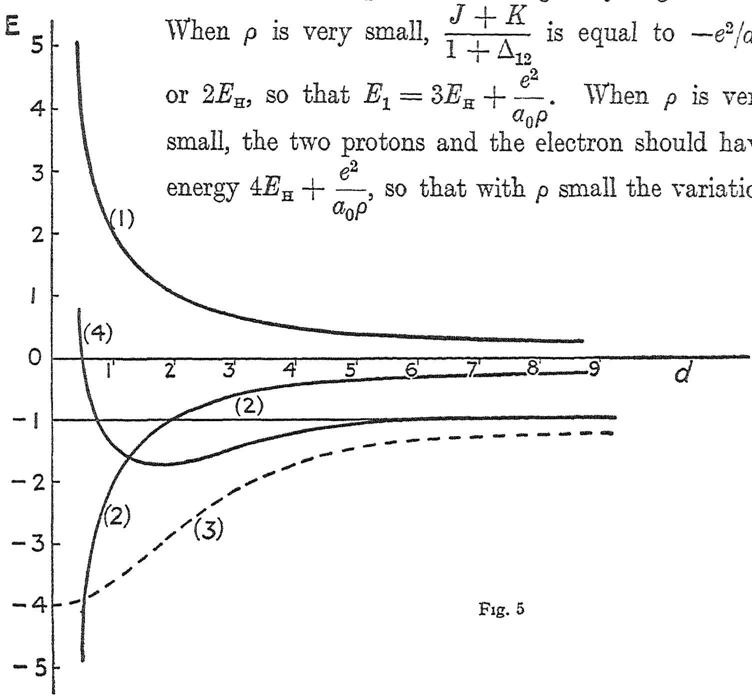


Fig. 5

method does not give the correct result. However E_1 has a minimum value near to $\rho = 2$ because $e^2/a_0\rho$ is positive and falls as ρ increases, while $\frac{J + K}{1 + \Delta_{12}}$ increases from $-e^2/a_0$ to 0 as ρ increases.

In fig. 5, curve (1) gives the values of e^2/d; curve (2) those of $-e^2/d$; curve (3) gives the values of $E_H + \frac{J + K}{1 + \Delta_{12}}$; and curve (4) those of E_1. The unit for d is a_0, and the unit of energy is $-E_H = e^2/2a_0$.

The H_2^+ ion is therefore stable with $\rho = 2$ and $E = E_1$. With $E = E_2$ there is no minimum, and so the ion is unstable.

The difference between E_1 and E_2 is nearly equal to twice $K = -\int u_1 \frac{e^2}{r_1} u_2 dv$. The variation of K with d produces an attraction

between the hydrogen atom and the proton and makes the ion stable with $E = E_1$. This energy K does not appear in classical theory and cannot be explained in classical terms. The integral K is called an exchange integral, and the attraction due to it is called an exchange force. Such forces are often very large and important. The name exchange for the integral is used because u_1 is for the electron near one proton, and u_2 for it near the other one, so that changing from u_1 to u_2 may be said to exchange the two positions.

27. The Hydrogen Molecule and the Helium Atom.

The hydrogen molecule consists of two protons and two electrons. The potential energy V is given by

$$V = \frac{e^2}{r_{AB}} - \frac{e^2}{r_{A1}} - \frac{e^2}{r_{A2}} - \frac{e^2}{r_{B1}} - \frac{e^2}{r_{B2}} + \frac{e^2}{r_{12}},$$

where the suffixes A and B denote the two protons, and the suffixes 1 and 2 the two electrons. r_{A1}, for example, is the distance from the proton A to the electron (1), and r_{12} the distance between the two electrons.

When the distance r_{AB} between the two protons is large, the molecule is merely two hydrogen atoms with lowest energy-state functions,

$$u_A(1) = (\pi a_0{}^3)^{-1/2}e^{-r_{A1}/a_0} \quad \text{and} \quad u_B(2) = (\pi a_0{}^3)^{-1/2}e^{-r_{B2}/a_0},$$

or $$u_A(2) = (\pi a_0{}^3)^{-1/2}e^{-r_{2A}/a_0} \quad \text{and} \quad u_B(1) = (\pi a_0{}^3)^{-1/2}e^{-r_{B1}/a_0},$$

according to whether the electron (1) is near the proton A and the electron (2) near B or vice versa.

We may, therefore, take

$$\phi = c_1 u_A(1)u_B(2) + c_2 u_A(2)u_B(1),$$

and make $\int \bar{\phi} H \phi dv$ as small as possible by varying c_1 and c_2, as with the hydrogen molecular ion.

The calculation is quite similar to that for the molecular ion, so only the results will be given here. It is found that either $c_1 = c_2$ or $c_1 = -c_2$. The energy with $c_1 = c_2$ is given by

$$E_1 = 2E_H + \frac{e^2}{r_{AB}} + \frac{2J + J' + 2K\Delta^{1/2} + K'}{1 + \Delta};$$

and with $c_1 = -c_2$,

$$E_2 = 2E_H + \frac{e^2}{r_{AB}} + \frac{2J + J' - 2K\Delta^{1/2} - K'}{1 - \Delta},$$

where $$J = -e^2 \int \frac{u_A{}^2(1)u_B{}^2(2)dv}{r_{B1}},$$

$$J' = -e^2 \int \frac{u_A(1)u_B(2)dv}{r_{12}},$$

$$K = -e^2 \int \frac{u_A(1)u_B(2)\,dv}{r_{A1}},$$

$$K' = -e^2 \int \frac{u_A(1)u_B(2)u_A(2)u_B(1)\,dv}{r_{12}},$$

$$\Delta = \int u_A(1)u_B(2)u_A(2)u_B(1)\,dv.$$

The energy E_1 is lower than E_2 mainly because of the negative value of K'. E_1 has a minimum value when r_{AB} is varied so that the molecule is stable when $E = E_1$ but not when $E = E_2$.

The integral K' is called an exchange integral because $u_A(1)u_B(2)$ is changed into $u_A(2)u_B(1)$ when the positions of the two electrons are exchanged. The attraction due to the variation of K' with r_{AB} is called an exchange force, and it makes the molecule stable. This exchange force cannot be explained classically, and has a mathematical rather than a physical character. The exchange force is much larger than the forces due to the variation of the integrals J, J' and K.

The helium atom consists of a nucleus with charge $2e$ and two electrons. The potential energy is given by

$$V = -\frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{e^2}{r_{12}},$$

where r_1 and r_2 are the distances between the nucleus and the electrons and r_{12} is the distance between the two electrons.

If there were only one electron, the lowest energy-state function would be $\frac{1}{\sqrt{\pi}}\left(\frac{2}{a_0}\right)^{3/2} e^{-2r_1/a_0}$. With two electrons, neglecting e^2/r_{12}, it would be $\frac{1}{\pi}\left(\frac{2}{a_0}\right)^3 e^{-2(r_1+r_2)/a_0}$. We may suppose that the repulsion between the two electrons would have an effect similar to that of reducing the nuclear charge $2e$ to a value between e and $2e$, or αe, say, so that the state function would be approximately $\phi = \frac{1}{\pi}\left(\frac{\alpha}{a_0}\right)^3 e^{-\alpha(r_1+r_2)/a_0}$. Regarding α as a variable parameter, an approximate value of the lowest energy E_1 can then be obtained by varying α so as to make $\int \bar{\phi} H \phi\, dv$ as small as possible.

The Hamiltonian operator is

$$H = -\frac{h^2}{8\pi^2 m}(\Delta_1 + \Delta_2) - \frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{e^2}{r_{12}},$$

where

$$\Delta_1 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2} \quad \text{and} \quad \Delta_2 = \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial y_2^2} + \frac{\partial^2}{\partial z_2^2},$$

x_1, y_1, z_1 being the co-ordinates of one electron and x_2, y_2, z_2 those of the other.

It is convenient to use $a_0 = h^2/4\pi^2me^2$, the radius of the smallest electron orbit in the hydrogen atom on Bohr's theory, as unit of length, and e^2/a_0 as unit energy. These units are called atomic units. To do this, replace r by ra_0, Δ by Δ/a_0^2 and H by $H\frac{a_0}{e^2}$, so that since $\frac{h^2}{8\pi^2m} = \frac{e^2a_0}{2}$, the equation for H becomes

$$H = -\tfrac{1}{2}(\Delta_1 + \Delta_2) - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_{12}}.$$

In these units $\phi = \frac{\alpha^3}{\pi}e^{-\alpha(r_1+r_2)}$ and is the lowest energy-state function for an atom with a nuclear charge α and two electrons which do not repel each other. The Hamiltonian H_0 for such an atom is given by

$$H_0 = -\tfrac{1}{2}(\Delta_1 + \Delta_2) - \frac{\alpha}{r_1} - \frac{\alpha}{r_2},$$

and the lowest energy is $2\alpha^2E_{\text{H}}$, where $E_{\text{H}} = -\frac{e^2}{2a_0}$ is the lowest energy of the hydrogen atom. With E_{H} in atomic units, $2\alpha^2E_{\text{H}} = -\alpha^2$, so that $H_0\phi = -\alpha^2\phi$. The equations for H and H_0 give

$$H = H_0 - (2 - \alpha)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) + \frac{1}{r_{12}}.$$

The integral $\int\bar{\phi}H\phi dv$ is therefore given by

$$\int\bar{\phi}H\phi dv = \int H_0\phi^2dv - (2 - \alpha)\int\phi^2\left(\frac{1}{r_1} + \frac{1}{r_2}\right)dv + \int\frac{\phi^2dv}{r_{12}}.$$

The first integral is equal to $-\alpha^2$, since $H_0\phi = -\alpha^2\phi$ and $\int\bar{\phi}\phi dv = 1$. The second integral is equal to 2α, and the third one to $\frac{5}{8}\alpha$, so that

$$\begin{aligned} E = \int\bar{\phi}H\phi dv &= -\alpha^2 - 2\alpha(2 - \alpha) + \tfrac{5}{8}\alpha \\ &= \alpha^2 - \tfrac{27}{8}\alpha. \end{aligned}$$

According to the variation method, the value of α can be found by putting $\frac{dE}{d\alpha} = 0$, which gives $2\alpha - \frac{27}{8} = 0$, or $\alpha = \frac{27}{16}$. With this value of α, $E = (\frac{27}{16})^2 - \frac{27}{8}\cdot\frac{27}{16} = -(\frac{27}{16})^2$ atomic units of energy. The atomic unit of energy is 27·2 electron volts, so that $E = -77{\cdot}5$ electron volts. The experimentally found value is -79, so the variation method in this case gives a result about two per cent too large.

28. Pauli's Exclusion Principle.

Consider an atom consisting of a nucleus and n electrons. Each electron is moving in the field due to the nucleus and the other $n - 1$

electrons. An approximate solution of the problem can be found by supposing that this field is constant, and a function of the distance from the nucleus only. An electron near the nucleus will have a field nearly equal to that of the nucleus, and an electron far from the nucleus will have a weaker field, due to the nucleus and the electrons nearer the nucleus. On such an approximate solution, the electrons may be regarded as independent systems.

Let the ith electron have energy E_i and state function $w_i(x_i)$ where x_i stands for all the co-ordinates of the ith electron. Also let the jth electron have energy E_j and state function $w_j(x_j)$. The state function for the two electrons will be $w_i(x_i)w_j(x_j)$ and the energy $E_i + E_j$.

Now suppose the two electrons exchange their co-ordinates so that the state function becomes $w_i(x_j)w_j(x_i)$. The two electrons are identical, so that the exchange should make no difference to the energy or to the chance of finding the electrons in $dv = dx_i dx_j$, where dx_i stands for the product of the differentials of all the co-ordinates of the ith electron and dx_j for the same thing for the jth electron.

However, the chance of finding the electrons in dv is not the same for $w_i(x_i)w_j(x_j)$ and $w_i(x_j)w_j(x_i)$, but if we take

$$w_{ij} = w_i(x_i)w_j(x_j) + w_i(x_j)w_j(x_i)$$

for the state function, then w_{ij} remains unchanged when the electrons exchange their positions, so this value of w_{ij} is satisfactory.

Also, if we take

$$w_{ij} = w_i(x_i)w_j(x_j) - w_i(x_j)w_j(x_i)$$

for the state function, then w_{ij} is changed to $-w_{ij}$ when the electrons exchange their positions. But $\bar{w}_{ij}w_{ij}$ remains unchanged when w_{ij} is changed to $-w_{ij}$, so this value of w_{ij} is also satisfactory.

The first value of w_{ij} is called a symmetrical function, and the second one an antisymmetrical function. Both values of w_{ij} leave the energy $E_i + E_j$ and $\bar{w}_{ij}w_{ij}$ unchanged when the electrons are exchanged.

To explain the spectra of atoms, it is necessary to assume that no two electrons in the atom can have the same state function. This was first pointed out by Pauli and is known as Pauli's exclusion principle.

If w_{ij} is symmetrical, then if $w_i(x_i) = w_j(x_j)$, $w_{ij} = 2w_i(x_i)w_i(x_j)$; but if w_{ij} is antisymmetrical and $w_i(x_i) = w_j(x_j)$, then

$$w_{ij} = w_i(x_i)w_i(x_j) - w_i(x_j)w_i(x_i),$$

so that $w_{ij} = 0$. This means that $\bar{w}_{ij}w_{ij} = 0$, so that there would be no electrons anywhere if $w_i(x_i) = w_j(x_j)$.

To explain Pauli's exclusion principle, it is therefore necessary to suppose that the state functions for electrons in atoms are always antisymmetrical. Note that this means that each pair of electrons must have an antisymmetrical state function.

Each electron has four co-ordinates, namely, x, y, z and a spin co-ordinate. The angular momentum or spin component along any direction may be either $-\frac{h}{4\pi}$ or $-\frac{h}{4\pi}$, so the spin co-ordinate has just two possible values. If the state function for an electron is $w(x, y, z, s)$ where s is the spin co-ordinate, then two electrons can have state functions $u(x, y, z)$ one with spin $\frac{h}{4\pi}$ and the other with spin $-\frac{h}{4\pi}$.

Each state of an atom with quantum numbers n, l and m can have two electrons in it with opposite spins.

It is supposed that the exclusion principle also applies to protons and neutrons. Photons, however, appear to have symmetrical state functions so that any number of photons can have the same state function.

29. Perturbation Theory.

If the potential $V(x, y, z)$ of an atomic system is changed, by the addition of a small additional potential $v(x, y, z)$, to $V + v$, the energy proper values and proper functions will also be changed. The small potential $v(x, y, z)$ may be supposed to be due to the action of some outside system on the atom and is usually called a perturbation. First suppose that the atom is not degenerate, so that there is only one proper function w_n for each energy E_n.

Let Schrödinger's equation for the unperturbed atom be

$$\Delta w_n + \frac{8\pi^2 m}{h^2}(E_n - V)w_n = 0.$$

For the perturbed system this becomes

$$\Delta w_n' + \frac{8\pi^2 m}{h^2}(E_n' - V - v)w_n' = 0.$$

Let $w_n' = w_n + u_n$ and $E_n' = E_n + \epsilon_n$, where u_n and ϵ_n are the small changes due to the perturbation. Putting $8\pi^2 m/h^2 = k$, this gives

$$\Delta w_n + \Delta u_n + k(E_n + \epsilon_n - V - v)(w_n + u_n) = 0$$

or, neglecting products of small quantities,

$$\Delta u_n + k(E_n u_n + \epsilon_n w_n - V u_n - v w_n) = 0.$$

Now any function of x, y, z may be expanded in a series of the orthogonal functions w_n, so we may put $u_n = \sum_m a_m w_m$ so that

$$\sum_m a_m \Delta w_m + k(E_n \sum_m a_m w_m - V \sum_m a_m w_m + \epsilon_n w_n - v w_n) = 0.$$

But $\Delta w_m - kVw_m = -kE_m w_m$, so that

$$k(E_n \Sigma a_m w_m - \sum_m a_m E_m w_m + \epsilon_n w_n - v w_n) = 0.$$

Multiply this by $\bar{w}_n$ and integrate, so that since $\int \bar{w}_n w_m dv = 0$ and $\int \bar{w}_n w_n dv = 1$, we get $\epsilon_n = \int \bar{w}_n v w_n dv$. The changes in the energy proper values are therefore equal to the diagonal terms in the matrix for v or v_{nn} (p. 116) when v is very small, or we may say that ϵ_n is equal to the average value of the perturbing energy v.

If the atom is degenerate, then we may have more than one proper function belonging to the energy E_n. Let these proper functions be $w_{n1}, w_{n2}, w_{n3}, \ldots$. The proper function w_n is then a linear combination of these, so that $w_n = \sum_k c_k w_{nk}$. We assume that u_n, the small change in w_n due to the perturbation v, can be expanded in a series of the proper functions like w_{nk}, so that $u_n = \sum_{ml} A_{ml} w_{ml}$. The equation

$$\Delta u_n + k(E_n - V)u_n = k(v - \epsilon_n)w_n$$

then gives

$$\sum_{ml} A_{ml}(\Delta + k(E_n - V))w_{ml} = k(v - \epsilon_n)\sum_k c_k w_{nk}$$

or

$$\sum_{ml} A_{ml}(E_n - E_m)w_{ml} = (v - \epsilon_n)\sum_k c_k w_{nk}.$$

Multiplying this by $\bar{w}_{pl}$ and integrating, we get, since $\int \bar{w}_{pl} w_{ml} dv = 0$ when $m \neq p$,

$$A_{pl}(E_n - E_p)\int \bar{w}_{pl} w_{pl} dv = \sum_k c_k \int \bar{w}_{pl}(v - \epsilon_n)w_{nk} dv.$$

Now let $p = n$ so that $\sum_k c_k \int \bar{w}_{nl}(v - \epsilon_n)w_{nk} dv = 0$ with $l = 1, 2, 3, \ldots$. From

$$\int \bar{w}_{nl} w_{nk} dv = 0, \quad l \neq k,$$
$$= 1, \quad l = k,$$

putting

$$v_{lk} = \int \bar{w}_{nl} v w_{nk} dv,$$

we get

$$\sum_k c_k v_{lk} = c_l \epsilon_n, \quad l = 1, 2, 3, \ldots$$

or

$$c_1(v_{11} - \epsilon_n) + c_2 v_{12} + c_3 v_{13} + \ldots = 0,$$
$$c_1 v_{21} + c_2(v_{22} - \epsilon_n) + c_3 v_{23} + \ldots = 0,$$
$$c_1 v_{31} + c_2 v_{32} + c_3(v_{33} - \epsilon_n) + \ldots = 0,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$

Eliminating the c's, this gives the determinantal equation

$$\begin{vmatrix} v_{11}-\epsilon_n & v_{12} & v_{13}\ldots \\ v_{21} & v_{22}-\epsilon_n & v_{23}\ldots \\ v_{31} & v_{32} & v_{33}-\epsilon_n\ldots \\ \ldots & \ldots & \ldots \end{vmatrix} = 0$$

to determine the values of ϵ_n. We may say that the values of ϵ_n are the proper values of the matrix v_{lk}, which is the part of the complete v matrix containing only proper functions belonging to the energy proper value E_n. The equations may also be solved for the coefficients c_k in the expansion $w_n = \sum_k c_k w_k$. It appears, therefore, that the perturbation determines these coefficients, which are therefore different for different perturbations.

As an example, consider the case of a two-dimensional oscillator with Hamiltonian

$$H = \frac{p_x^2}{2m} + \frac{p_y^2}{2m} + \tfrac{1}{2}\mu x^2 + \tfrac{1}{2}\mu y^2$$

and Schrödinger equation

$$\frac{\partial^2 w_0}{\partial x^2} + \frac{\partial^2 w_0}{\partial y^2} + \frac{8\pi^2 m}{h^2}(E - \tfrac{1}{2}\mu(x^2+y^2))w_0 = 0.$$

We may put $w_0 = w_0(x)w_0(y)$ where $w_0(x)$ is a solution of

$$\frac{\partial^2 w_0(x)}{\partial x^2} + \frac{8\pi^2 m}{h^2}(E_x - \tfrac{1}{2}\mu x^2)w_0(x) = 0$$

and $w_0(y)$ a solution of

$$\frac{\partial^2 w_0(y)}{\partial y^2} + \frac{8\pi^2 m}{h^2}(E_y - \tfrac{1}{2}\mu y^2)w_0(y) = 0$$

and $E = E_x + E_y$.

We have (see § 19, p. 103) $E_x = h\nu(n_x + \frac{1}{2})$ and $E_y = h\nu(n_y + \frac{1}{2})$, so that

$$E = h\nu(n_x + n_y + 1).$$

The proper functions with $E = h\nu$, so that $n_x = 0$ and $n_y = 0$, are $w_0(x) = e^{-ax^2}$, $w_0(y) = e^{-ay^2}$, omitting the constant factors, so that $w_0(x, y) = e^{-ar^2}$.

With $E = 2h\nu$ so that either $n_x = 1$, $n_y = 0$ or $n_x = 0$, $n_y = 1$, the proper functions are

$$w_0(x) = xe^{-ax^2} \quad \text{or} \quad w_0(x) = e^{-ax^2}$$

and $$w_0(y) = e^{-ay^2} \quad \text{or} \quad w_0(y) = ye^{-ay^2},$$

so that $$w_0(x, y) = xe^{-ar^2} \quad \text{or} \quad w_0(x, y) = ye^{-ar^2}.$$

It appears that with $E = 2h\nu$ the system has two proper functions and so is degenerate. Any linear combination of the two proper functions is also a solution of Schrödinger's equation, so

$$w_0 = c_1 x e^{-ar^2} + c_2 y e^{-ar^2}$$

is a solution. Now suppose that a small force acts on the particle, giving it a small additional potential energy $f(x, y)$. We may regard $f(x, y)$ as a perturbation and so have $\sum_k c_k v_{lk} = c_l \epsilon$, where $v = f(x, y)$ and ϵ is the change in energy due to the perturbation. Thus we have with $l = 1$

$$c_1 v_{11} + c_2 v_{12} = c_1 \epsilon,$$

and with $l = 2$

$$c_1 v_{21} + c_2 v_{22} = c_2 \epsilon.$$

The matrix elements are

$$v_{11} = \int x^2 v e^{-2ar^2} dx dy,$$

$$v_{22} = \int y^2 v e^{-2ra^2} dx dy,$$

and

$$v_{12} = v_{21} = \int xyve^{-2ar^2} dx dy,$$

omitting the normalizing factors. The values of ϵ are given by

$$\begin{vmatrix} v_{11} - \epsilon & v_{12} \\ v_{21} & v_{22} - \epsilon \end{vmatrix} = 0$$

or $(v_{11} - \epsilon)(v_{22} - \epsilon) = v_{12}^2$, so the values of c_1, c_2 and ϵ can be easily calculated.

30. Stark Effect.

If an atom is in an electric field F, along the x-axis, the potential of an electron in the atom is increased by Fex. For example, Schrödinger's equation for an hydrogen atom becomes

$$\Delta w + \frac{8\pi^2 m}{h^2}\left(E + \frac{e^2}{r} - Fex\right)w = 0,$$

the additional potential Fex can be regarded as a perturbation, and so the changes in the proper values of the energy due to the electric field can be calculated. The frequencies ν of the spectral lines emitted are given by $E_n - E_m = h\nu$, so the electric field alters the frequencies. The atoms are degenerate, so that each line is split up into several lines. This effect was discovered by Stark, and the observed changes of frequency are found to agree well with those calculated.

31. Angular Momentum.

In classical dynamics the z component of the angular momentum of an electron about the z-axis, or m_z say, is equal to $xp_y - yp_x$, and

it is defined by the same expression in quantum mechanics. It is therefore the operator $\frac{h}{2\pi i}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)$, so that

$$m_z w = \frac{h}{2\pi i}\left(x\frac{\partial w}{\partial y} - y\frac{\partial w}{\partial x}\right).$$

Putting $x = r\cos\phi$ and $y = r\sin\phi$, this gives $m_z w = \frac{h}{2\pi i}\frac{\partial w}{\partial \phi}$. We may therefore put $w = fe^{2\pi i m_z'\phi/h}$, where m_z' is a proper value of m_z and f does not involve ϕ. Now w must be a single valued function of ϕ, so that we must have $m_z' = \frac{mh}{2\pi}$, where $m = 0, \pm 1, \pm 2, \ldots$, since then $w = fe^{im\phi}$, which is unchanged when ϕ increases by 2π. The integer m is the same as the quantum number m which appeared in the solution of Schrödinger's equation for central forces.

The difference $m_z zw - zm_z w$ is equal to

$$\frac{h}{2\pi i}\frac{\partial}{\partial \phi}(zw) - z\frac{h}{2\pi i}\frac{\partial w}{\partial \phi} = 0,$$

so that $m_z z - zm_z = 0$. The matrix element $(m_z z - zm_z)_{nm}$ is equal to

$$\sum_k\left\{(m_z)_{nk}z_{km} - z_{nk}(m_z)_{km}\right\}.$$

But the proper values of m_z are constants equal to $nh/2\pi$, so that the m_z matrix is diagonal and $(m_z)_{nk} = 0$ unless $n = k$. Hence

$$(m_z z - zm_z)_{nm} = m_z' z_{nm} - z_{nm} m_z'' = 0,$$

where m_z' and m_z'' are two proper values of m_z. We have therefore $(m_z' - m_z'')z_{nm} = 0$. This shows that $z_{nm} = 0$ unless $m_z' = m_z''$. Now z_{nm} is proportional to the matrix element of the z component of the electric moment due to the displacement of the electron from the origin, so that if $z_{nm} = 0$ there will be no radiation due to this component. This means that if a transition occurs resulting in the emission of a photon due to an electric oscillation along the z-axis, then the angular momentum about the z-axis is unchanged by the transition. This result is called the *selection rule* for m_z for oscillations along the z-axis.

32. Selection Rules.

We can get the selection rule for m_z for oscillations along the x- and y-axes in a similar way.

We have

$$m_z xw = \frac{h}{2\pi i}\frac{\partial}{\partial \theta}(xw) = \frac{h}{2\pi i}\left(\frac{\partial x}{\partial \theta}w + x\frac{\partial w}{\partial \theta}\right).$$

As an example of transitions due to a perturbation we will take the case of transitions due to light. Let the atom be in a beam of parallel plane polarized light going along the x-axis, and let the electric field F in the light be parallel to the y-axis. The potential energy of an electron in the atom due to the light is then $v = Fey$. If the wave-length of the light is large compared with the dimensions of the atom, F will be constant over the volume of the atom, so that

$$v_{mn}^* = eFy_{mn}^* = eFy_{0mn}e^{-2\pi i(E_n - E_m)t/h}$$

and
$$\bar{c}_{0m}c_{0m} = \frac{4\pi^2 e^2}{h^2}\,|\,y_{0mn}\,|^2 \left| \int_0^{t'} Fe^{-2\pi i(E_n - E_m)t/h}\,dt \right|^2.$$

The electric field $F(t)$ during the time interval 0 to t' may be represented by a Fourier's series,†

$$F(t) = \sum_{-\infty}^{+\infty} c_n e^{2\pi i n t/t'},$$

where
$$c_n = \frac{1}{t'}\int_0^{t'} F(t)e^{-2\pi i n t/t'}\,dt,$$

and the average value of the square of the component with frequency $\nu = n/t'$ is $2\,|\,c_n\,|^2$. The average energy per cm.^{3} in the radiation for the component of frequency ν is therefore $\frac{1}{2\pi}\,|\,c_n\,|^2$, so that the energy per cm.^{2} in time 0 to t' is $\frac{ct'}{2\pi}\,|\,c_n\,|^2$. The number of components with frequencies between ν and $\nu + \delta\nu$ is δn, given by $\delta\nu = \delta n/t'$, so that the number of components per unit range of frequency ($\delta\nu = 1$) is equal to t'. The energy per cm.^{2} in time 0 to t' per unit range of frequency at the frequency ν is therefore given by

$$E_\nu = \frac{ct'^2}{2\pi}\,|\,c_n\,|^2.$$

But
$$c_n = \frac{1}{t'}\int_0^{t'} F(t)e^{-2\pi i n t/t'}\,dt,$$

so that
$$E_\nu = \frac{c}{2\pi}\left| \int_0^{t'} F(t)e^{-2\pi i\nu t}\,dt \right|^2,$$

where $\nu = n/t'$.

We have therefore

$$c_{0m}c_{0m} = \frac{8\pi^3 e^2}{ch^2}\,|\,y_{0mn}\,|^2 E_\nu \quad \text{with} \quad \nu = \frac{E_n - E_m}{h}.$$

† See p. 415.

For a steady flow of light energy, the energy E_ν will be proportional to the interval 0 to t', provided 0 to t' is long compared with the time of one light vibration or $1/\nu$. The chance of a transition from w_n to w_m or w_m to w_n during the interval is then proportional to the length of the interval as we should expect.

The chance of a transition from energy E_n to E_m is determined by the energy per unit range of frequency in the light at the frequency $(E_m - E_n)/h$, so if E_m is greater than E_n, we may suppose the atom absorbs a photon of energy $h\nu = E_m - E_n$. If the atom is initially in the state with the higher energy E_m, the transition results in an emission of energy $E_m - E_n$ as a photon of energy $h\nu$. It may be possible for the photon causing the transition to receive the emitted energy, which would change it into a photon of energy $2h\nu$.

35. **Collisions.**

If the path of a free particle such as a proton or electron lies very close to an atom, there will be an interaction between them which will change the direction of the particle and may cause one to gain energy at the expense of the other. If the energy of the particle before the collision is ϵ_1 and that of the atom E_n, and the energies after the collision are ϵ_2 and E_k, then $\epsilon_1 + E_n = \epsilon_2 + E_k$, provided there is no loss of energy by radiation or otherwise. If the particle is an electron or a proton, we may suppose the atom at rest before and after the collision, provided the mass of the atom is large compared with that of the particle.

The proper function of a particle before the collision may be taken to be $ae^{2\pi i p_x x/h}$, omitting the time factor and supposing that it is moving parallel to the x-axis with momentum p_x. Let $2\pi p_x/h = k_0$ so that $ae^{2\pi i p_x x/h} = ae^{ik_0 x}$. This makes the chance of finding the particle in an element of volume dv equal to $a^2 dv$. But this chance must be zero since the particle may be anywhere, so that $a^2 = 0$. To avoid this difficulty we suppose there is one particle per unit volume throughout space all moving in the same direction with the same velocity. The chance of finding a particle in $dx\,dy\,dz$ is then just equal to dv, so that $a^2 = 1$. We may also suppose that the atom is at the origin and that whenever, as a result of a collision, it makes a transition it is immediately replaced by an atom with the original energy E_n. The number of transitions per unit time from E_n to any given E_k with the particle going off after the collision inside a small cone with its vertex at the atom and its axis making an angle θ with the x-axis will then be constant. We assume that the distribution is symmetrical about the x-axis. Let the number of particles per unit volume moving along the small cone be $N_k(\theta)$. The number per unit time is then $N_k(\theta) r^2 p_k d\omega/m$, where p_k/m is the velocity of the particles along the cone and $d\omega$ the solid angle of the cone. The number of

particles, before collisions, falling on an area a per unit time is ap_x/m, so if we put

$$N_k(\theta)r^2p_k d\omega/m = ap_x d\omega/m$$

so that $a = \frac{p_k}{p_x} r^2 N_k(\theta)$, then a is the target area of the atom per unit solid angle for collisions in which the particle goes off in a direction θ and E_n is changed to E_k. The proper function for the particles in the direction θ, for large r, may be taken to be $r^{-1}u(k, \theta)e^{2\pi i p_k r/h}$ or $r^{-1}u(k, \theta)e^{ikr}$, so that $N_k(\theta) = r^{-2}|u(k, \theta)|^2$ and $a = \frac{p_k}{p_x}|u(k, \theta)|^2$. The problem then is to find $u(k, \theta)$.

Let Schrodinger's equation for the particles before they are scattered be $H_p\phi = E_p\phi$, and for the atom $H_a w_n = E_n w_n$, where w_n is a function of the co-ordinates of the electrons in the atom and E_n one of its energy proper values. The equation for the particles and the atom will then be

$$(H_p + H_a + V)\psi = E\psi,$$

where V is the interaction energy. It is a function of the co-ordinates of the atomic electrons and of the colliding particle. We assume that V is small, so that the proper function of the incident particles and the proper functions of the atom are not appreciably altered by the interaction. The function ψ will be equal to $\phi w_n + u$, where u belongs to the particles deflected by the collisions. We have therefore

$$(H_p + H_a + V)(\phi w_n + u) = E(\phi w_n + u),$$

which, neglecting the product Vu, gives

$$(E - H_p - H_a)u = V\phi w_n.$$

We may replace $E - H_a$ by the energy of a particle or $p^2/2m$. Putting

$$2\pi p/h = k \quad \text{and} \quad H_p = -\frac{h^2}{8\pi^2 m}\Delta,$$

where

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2},$$

we get

$$(\Delta + k^2)u = \frac{8\pi^2 m}{h^2} V\phi w_n.$$

Now let $u = \Sigma\phi_k w_k$, where ϕ_k is a proper function of the scattered particles, so that

$$\Sigma(\Delta + k^2)\phi_k w_k = \frac{8\pi^2 m}{h^2} V\phi w_n.$$

Multiplying by $\bar{w}_k$ and integrating with respect to the co-ordinates of the atomic electrons only, we get

$$(\Delta + k^2)\phi_k = \frac{8\pi^2 m}{h^2}\phi\int \bar{w}_k V w_n dv'.$$

The solution of this equation is similar to that of the wave equation $\Delta\psi - \frac{1}{u^2}\ddot{\psi} = \omega$, for if $\omega = \omega_0 e^{-2\pi i\nu t}$ and $\psi = \psi_0 e^{-2\pi i\nu t}$, the wave equation becomes $(\Delta + k^2)\psi = \omega$. The solution of the wave equation shows that ψ, at time t, at any point P is the sum of disturbances $-\frac{[\omega]dv}{4\pi R}$, one for each element of volume, where $[\omega]$ is the value of ω at the time $t - R/u$ and R is the distance from dv to P. If

$$\psi = \psi_0 e^{-2\pi i\nu t} \quad \text{and} \quad \omega = \omega_0 e^{-2\pi i\nu t},$$

then we get the sum of effects

$$-\frac{\omega_0 e^{-2\pi i\nu(t-R/u)}}{4\pi R} = -\frac{\omega_0}{4\pi R}e^{-2\pi iEt/h+ikR},$$

where $k = \frac{2\pi mv}{h} = 2\pi/\lambda$. Omitting the time factor, we get

$$\psi_0 = -\int\frac{\omega_0 e^{+ikR}}{4\pi R}dv.$$

We may therefore suppose that ϕ_k is the sum of effects equal to

$$-\frac{2\pi m}{h^2}\phi\frac{e^{+ikR}}{R}dv\int\bar{w}_k V w_n dv'$$

for each element of volume $dv = dx\,dy\,dz$. The element dv' is equal to $dx'\,dy'\,dz'\,dx''\,dy''\,dz'' \ldots$, where $x', x'', \ldots$ are the co-ordinates of the atomic electrons. The effects may be regarded as due to each element of volume dv scattering the incident de Broglie waves ϕ. We have therefore for ϕ_k at a point P, putting $\int\bar{w}_k V w_n dv' = V_{kn}$ and $\phi = e^{ik_0 x}$,

$$\phi_k = -\frac{2\pi m}{h^2}\int\frac{e^{+ikR+ik_0 x}}{R}V_{kn}dv.$$

V_{kn} is a function of the co-ordinates of the colliding particle. If the atom is symmetrical, we may take it to be a function of the distance r, say, of the particle from the origin. The momentum of the particles before a collision is given by $k_0 = 2\pi p_x/h$, and after a collision by $k = 2\pi p_k/h$. The momentum p_k is directed towards the point P. The value of p_k is determined by $E_n + \frac{p_x^2}{2m} = E_k + \frac{p_k^2}{2m}$. w_n is the

proper function of the atom before a collision, and w_k that after one. $\bar{w}_k V w_n$ will be small when r is large, so if R is large we may put $R = R_0 - r\cos\phi$, where $R_0 = OP$ and ϕ is the angle between OP and r. We have then

$$\phi_k = -\frac{2\pi m}{h^2 R_0} e^{ikR_0} \int e^{-ikr\cos\phi + ik_0 x} V_{kn}\, dv.$$

We see that ϕ_k is the same thing as $r^{-1}u(k, \theta)e^{ikr}$ above, so that the target area α is given by

$$\alpha = \frac{p_k}{p_v}\frac{4\pi^2 m^2}{h^4}\left|\int e^{-ikr\cos\phi + ik_0 x} V_{kn}\, dv\right|^2.$$

36. **Elastic Collisions.**

If the collisions are elastic, so that the atomic energy is unchanged, then $k_0 = k$ and $w_k = w_n$, so that

$$\alpha = \frac{4\pi^2 m^2}{h^4}\left|\int e^{ik(x - r\cos\phi)} V_{nn}\, dv\right|^2.$$

Let OP lie in the xy-plane and consider a plane through Oz which bisects the angle θ between Ox and OP. A parallel beam of light going along the x-axis would be reflected by this plane or any parallel plane along OP, so that all the light rays near to the x-axis would have paths to P of equal length. Now $x + R = x + R_0 - r\cos\phi$ is the length of such a path, and therefore $x - r\cos\phi$ is constant over any of the parallel planes. Let y be the length of the perpendicular from the origin on to one of the parallel planes. Then $x - r\cos\phi$ is equal to $-2y\sin(\theta/2)$, so that the integral becomes

$$\int e^{-i\mu y} V_n\, dv, \quad \text{where} \quad \mu = 2k\sin(\theta/2).$$

Changing to new polar co-ordinates with $y = r\cos\psi$ and $dv = 2\pi r^2 \sin\psi\, d\psi\, dr$, we find, assuming V_n to be a function of r only, and since

$$\int e^{-i\mu r\cos\psi}\, 2\pi r^2 \sin\psi\, d\psi\, dr = 4\pi\int_0^\infty \frac{\sin\mu r}{\mu r} r^2\, dr,$$

$$\alpha = \frac{64\pi^4 m^2}{h^4}\left|\int_0^\infty \frac{\sin\mu r}{\mu r} V_n r^2\, dr\right|^2.$$

As a simple example take the case of alpha rays of mass m and charge $2e$ scattered by atoms with a nuclear charge Ze. The alpha particle is not appreciably deflected until it gets very near the nucleus, where its potential energy is $2Ze^2/r$. The interaction of the alpha ray

with the electrons has no appreciable effect on the alpha ray and so may be neglected. We have therefore

$$V_n = \frac{2Ze^2}{r}\int \bar{\psi}_n \psi_n d\tau = 2Ze^2/r.$$

It is convenient to take* $V_n = \frac{2Ze^2}{r} e^{-\beta r}$ and then let β approach zero. We have then, since

$$\int_0^\infty \sin \mu r \, . \, e^{-\beta r} dr = \frac{\mu}{\mu^2 + \beta^2}$$

or $1/\mu$ when β is very small,

$$\alpha = \frac{64\pi^4 m^2 \, . \, 4Z^2 e^4}{h^4 \mu^4}.$$

Putting $\mu = 2k \sin(\theta/2)$ and $k = 2\pi mv/h$, where v is the velocity of the particles, we get

$$\alpha = \frac{Z^2 e^4}{m^2 v^4} \operatorname{cosec}^4(\theta/2).$$

This agrees with the classical theory of the single scattering of alpha rays. α is the target area, per unit solid angle, for rays scattered in the direction making an angle θ with that of the incident rays.

Several other applications of quantum mechanics are considered in other chapters. The quantum theory of free electrons in metals, of line spectra and of nuclear disintegrations may be mentioned as specially interesting.

* Any function of r, which is equal to unity when r is small and drops gradually to zero as r increases, may be used instead of $e^{-\beta r}$.

REFERENCES

1 *Wave Mechanics.* Mott

2. *Elementary Quantum Mechanics.* Gurney.
3. *Quantum Mechanics.* Condon and Morse.
4. *Quantum Mechanics.* Dirac.
5. *The Mysteries of the Atom.* H. A. Wilson.
6. *Quantum Chemistry* Eyring, Walter, Kimball.
7. *Mathematics of Physics and Chemistry.* Margenan and Murphy.

CHAPTER VI

The Critical Potentials of Atoms

1. Quantum Theory and Critical Potentials.

According to the quantum theory atoms can only exist in a series of states characterized by definite energies $W_1, W_2, W_3 \ldots W_n$ . . In these so-called stationary states the atom does not emit radiation. The normal state of the atom is the state of smallest energy W_1. To change an atom from its normal state to another state requires energy $W_n - W_1$ to be given to the atom. An atom not in the normal state is usually called an excited atom. Excited atoms emit radiation of frequency ν given by $h\nu = W_m - W_n$ when they change from a state having energy W_m to one having energy W_n, and they usually quickly revert to the normal state.

If a stream of electrons all moving with the same velocity v is passed into a gas it is found that the gas atoms may be excited by collisions with the electrons, provided the kinetic energy of the electrons is great enough. The kinetic energy $\frac{1}{2}mv^2$ must be at least equal to $W_n - W_1$ for the electrons to be able to change an atom from its normal state with energy W_1 to the state having energy W_n.

If P denotes the potential difference required to give an electron the kinetic energy $\frac{1}{2}mv^2$ so that $Pe = \frac{1}{2}mv^2$, where e is the electronic charge, then if $Pe = W_m - W_n$ the potential difference P is called a *critical potential* of the atom for which W_m and W_n are possible energies. The potential difference required to give an electron just enough energy to ionize the atom, that is to knock an electron right out of it, is called an ionization potential of the atom. If W' is the energy of the ionized atom, and W_1 its normal energy, then $Pe = W' - W_1$ gives the ionization potential of the atom when in its normal state.

2. Lenard's Measurements.

The critical potentials of atoms were first directly measured by Lenard with an apparatus shown diagrammatically in fig. 1. Electrons from a hot filament F are attracted by a wire grating or grid G which is charged positively by connecting it to a battery, and some of them pass through the grating. A plate P on the other side of G is connected

to one pair of the quadrants of a quadrant electrometer E and the other pair of quadrants is connected to the earth. The plate P can be insulated when desired by opening the key K. If P_G denotes the potential of the grid and P_F that of the filament, then $P_G - P_F = V_a$ is called the accelerating potential, and the kinetic energy of the electrons from the filament, when they reach the grid, is approximately given by $\frac{1}{2}mv^2 = V_a e$. The potential difference between the grid and the plate P is called the retarding potential $V_r = P_G$. The filament is kept at a positive potential less than P_G so that V_r is greater than V_a. The filament, grid, and plate are inside a bulb connected to a pump through a tube T, and the gas pressure in the bulb is adjusted so that the mean free path of the electrons is greater than the distance from the filament to the grid. The electrons which pass through the grid are repelled by the plate and, since $V_r > V_a$, none of them can get to it. If, however, the energy of the electrons is great enough they ionize some of the gas molecules by collisions in the space between the grid and plate, and the positive ions produced are attracted by the plate P and give it a positive charge which is indicated by the electrometer when the key K is open. If V_a is gradually increased, keeping V_r always greater than V_a, then the smallest value of V_a for which the plate gets any positive charge can be found.

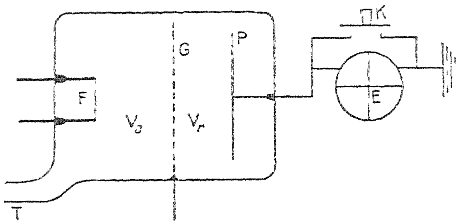


Fig. 1

Fig. 2 shows the relation between V_a and the positive charge received by the plate in a given time with hydrogen gas in the apparatus. No charge is observed until V_a is about 10 volts. The charge increases slowly up to about 16 volts and then begins to increase rapidly with the accelerating potential. This shows that hydrogen has critical potentials of 10 and 16 volts approximately.

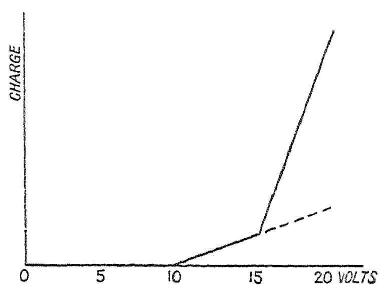


Fig. 2

The positive charge received by the plate P, however, is not necessarily due to positive ions; it may be caused by ultra-violet light emitted by the gas falling on the plate and causing it to emit electrons. The electrons from the filament may have enough energy to excite the gas molecules but not enough to ionize them. If the molecules

are excited by the electron collisions they emit light when they go back to the normal state, and this light may cause the plate to emit electrons and so become positively charged. Lenard's method shows that the hydrogen has critical potentials at about 10 and 16 volts, but it does not enable us to determine whether these potentials are ionization potentials or excitation potentials.

3 Results of Franck and Hertz with Mercury Vapour.

A great many modifications of Lenard's original method have been devised but only a few of these can be described here A full account is given in the *Bulletin of the National Research Council*, No 48, "Critical Potentials", by K T. Compton and F L Mohler

If the gas pressure in the apparatus is increased so that the mean free path of the electrons becomes considerably smaller than the distance from the filament to the grid and if also the retarding potential is made quite small (about ½ volt), then, with monatomic gases, the current to the plate rises and falls periodically as the accelerating potential is increased. Results obtained in this way by Franck and Hertz with mercury vapour are shown in fig 3.

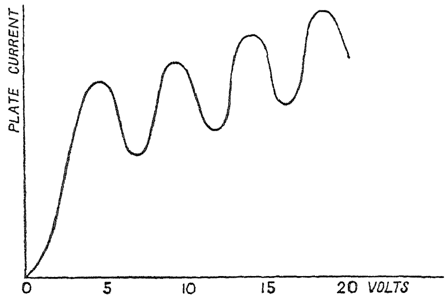


Fig 3

The potential differences between the successive maxima are 4 9 volts. It is supposed that the electrons, when their kinetic energy corresponds to less than 4 9 volts, collide with the mercury atoms without loss of energy. The collisions are then said to be elastic. An electron from the filament therefore loses no energy due to collisions as it moves towards the grid until its energy corresponds to 4·9 volts, and then it excites a mercury atom and loses all its kinetic energy. The energy of the electron then increases again as it moves along in the electric field until it again corresponds to 4 9 volts, when it again excites a mercury atom and loses its energy, and so on. In order to get to the plate an electron must arrive at the grid with energy greater than that corresponding to the small retarding potential V_r. Hence as V_a is gradually increased the current increases until V_a is about 4·9 volts, when it drops, because the electrons lose their energy just before reaching the grid. The current then increases again as V_a is increased until it reaches $2 \times 4{\cdot}9$ volts, when it again falls and so on. The lowest critical potential of mercury atoms is therefore 4 9 volts. Similar results have been obtained with helium and other monatomic gases.

More precise results can be obtained by using two grids, one near the filament and the other near the plate. The one near the plate is kept at a slightly higher potential (about 0 1 volt) than the one near the filament. The electrons which pass through the first grid then move with nearly constant velocity in the space between the two grids, so that the number of collisions occurring with any given energy is greatly increased, and the breaks in the curve giving the relation between plate current and accelerating potential are much more distinct. The second grid can be used with Lenard's method at low pressures and with Franck and Hertz's method at higher pressures. In this way a great many critical potentials

have been observed; for example, in mercury there are critical potentials of 4·68, 4·9, 5·29, 5·78, and 6·73 volts.

4. Methods distinguishing Ionization and Excitation Potentials.

A modification which enables ionization and excitation potentials to be easily distinguished was used by K. T. Compton and by Boucher. Two plates were provided, one consisting of a metal sheet, as usual, and the other of a grating of fine wire. The two plates were connected together and either of them could be moved into position in front of the grids when desired. The area of the metal sheet was much greater than that of the grating so that the current due to radiation was much greater with the sheet than with the grating, but the current due to positive ions was about the same with either plate. The ratio of the plate current with the sheet to the plate current with the grating remained nearly constant as the accelerating potential was increased so long as only excitation of atoms was taking place, but began to fall rapidly as soon as ionization began.

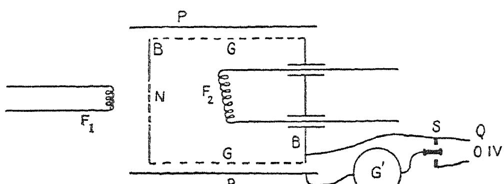


Fig. 4

Fig. 4 shows diagrammatically an apparatus due to Hertz which enables two different methods to be used. BB is a cylindrical box with flat ends. In one end at N is a hole covered with a grating. The sides of the box GG are made of wire gauze. Surrounding the box is a metal cylinder PP. A filament F_1 is set up near the hole at N and there is a second filament F_2 inside the box.

To determine ionization potentials the filament F_2 is heated and the current from box to filament due to a small potential difference is measured. The temperature of F_2 is raised until the current is independent of the temperature, showing that the current is limited by the space charge on the electrons emitted. If now electrons from the other filament F_1 are accelerated into the box through the hole at N by a potential difference V_a, they have no effect on the current unless they produce positive ions in the box by collisions with the gas molecules present. The positive ions neutralize some of the space charge and so increase the current. Owing to the mass of the positive ions being much greater than that of the electrons the ions remain in the space inside the box much longer than the electrons, so that the increase of current is enormously greater than the current carried by the positive ions. This is therefore a sensitive method for detecting an ionization potential.

The same apparatus can also be used to detect any kind of critical

potential by another method. The filament F_2 inside the box is not used but is merely connected to the box. Electrons from F_1 are accelerated into the box through N and some of them escape from the box through the gauze sides to the cylinder P. The current to P carried by these electrons is measured with a galvanometer G′ connected to P. A switch S is provided in the galvanometer circuit by means of which a potentiometer Q arranged to give a small P.D. of about 0·1 volt can be introduced into the circuit when desired. The current is observed with the small P D. on, and also with it off. With the P.D. on, the electrons from the box cannot get to P unless they have enough energy to carry them across the small P.D., but with the P.D. off all the electrons from the box get to P. The difference between the galvanometer currents with the P.D. off and with it on therefore gives the number of electrons coming out of the box which have kinetic energies between zero and the energy corresponding to the small P.D.

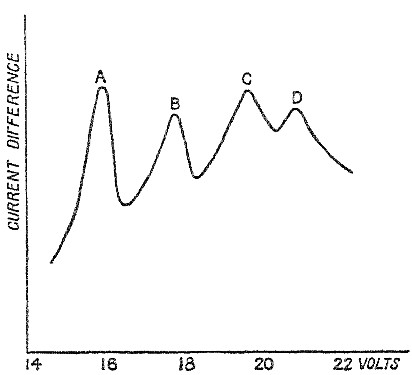


Fig 5

If V_a is the accelerating potential driving the electrons from F_1 into the box and if V_c is a critical potential for the gas in the box, then if $V_a - V_c$ is between zero and the value of the small P.D. there will be electrons in the box with energies between zero and that corresponding to the small P.D., some of which will escape from the box and produce a difference between the two currents observed. If, however, $V_a - V_c$ is greater than the small P.D. or less than zero there will be no such electrons present and the two currents will be equal.

As the accelerating potential is gradually increased the two currents remain equal until a critical potential is reached, when a difference appears which rises to a maximum and falls back to zero when $V_a - V_c$ becomes greater than the small P.D. On V_a being further increased, when a second critical potential is reached a difference between the two currents appears again and then disappears, and so on. Fig. 5 shows the results obtained by Hertz with a mixture of neon and helium. The two maxima at A and B are due to critical potentials of helium and those at C and D to neon, as can be shown by trying the two gases separately.

All the above methods are subject to an error on account of the initial velocity of the electrons emitted by the hot filament. The energy of the electrons is greater than that corresponding to the

accelerating potential V_a. This error is best determined by measuring known critical potentials, or by measuring the difference between the critical potential which it is desired to find and a known critical potential. Possible errors due to contact potential differences are also eliminated in this way.

5. Agreement of Results with Quantum Theory.

By means of the above and other more or less similar methods the critical potentials of many different atoms and molecules have been determined. These results provide a remarkable confirmation of one of the fundamental assumptions of the quantum theory, viz. that the energy of atoms does not vary continuously but that atoms can only exist in a series of definite states having definite energies. Electrons collide elastically with an atom unless they have enough kinetic energy to change the energy of the atom from its actual state to another possible state.

According to Bohr's quantum theory of spectra, moreover, we have

$$h\nu = W_m - W_n,$$

so that if P denotes the critical potential corresponding to the change from energy W_n to energy W_m, we have

$$h\nu = Pe.$$

If P is expressed in volts and ν as the wave number or number of waves in 1 cm. in a vacuum, then on substituting the values of h and e we get

$$P = 1{\cdot}2344 \times 10^{-4}\nu.$$

It is found that the observed critical potentials in many cases agree accurately with the values calculated by means of this equation. Since the wave number can be determined with much greater accuracy than the critical potential P, it follows that when the spectral line corresponding to a particular critical potential has been definitely identified, then the value of the critical potential can be most accurately obtained by calculation from the wave number of the spectral line.

The results which have been obtained with the alkali metals are especially simple. The vapours of these metals are monatomic, and are found to absorb the lines of the principal series in their spectra strongly. This shows that the lines of this series are emitted when the atoms revert to the normal state from a series of states having greater energies. See Chap. VIII.

An excitation potential has been found for these atoms which agrees with that calculated from the wave number of the first line in the principal series, and also an ionization potential which agrees with that calculated from the wave number of the limit of the principal

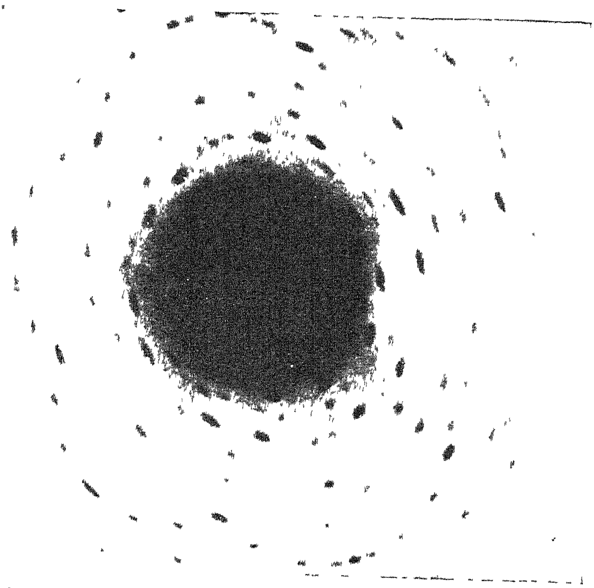


Fig. 8, Chap. VII.—Diffraction of X-Rays by a Crystal of Beryl
(From *X Rays and Crystal Structure*, Sir W. H. and W. L. Bragg)

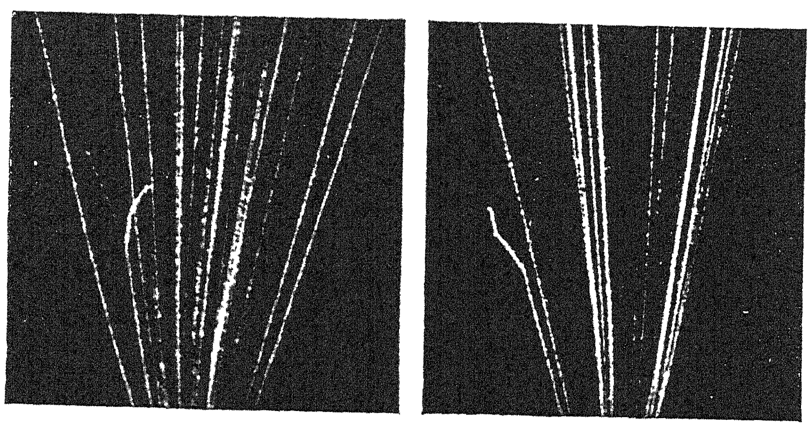


Fig. 5, Chap. XIII.—Photographs of α-ray tracks through nitrogen, by P. M. S. Blackett, showing collision between α ray and nitrogen atom, at which a track branches into a thin straight track due to the ejected proton and a thick track due to the nitrogen atom combined with the α-ray. See p. 280

The following table gives the observed and calculated results in volts.

Metal	Wave Number of Limit.	Ionization Potential		Wave Numbers of First Line	Excitation Potential	
		Cal.	Obs.		Cal.	Obs.
Sodium	41 449	5·116	5·13	16,973 16,956	2·095 2·093	2·12
Potassium	35,006	4·321	4·1	13,043 12,985	1·610 1·603	1·55
Rubidium	33,689	4·159	4·1	12,817 12,579	1·582 1·553	1·6
Caesium	31,405	3·877	3·9	11,732 11,178	1·448 1·380	1·48

The first lines of the principal series are doublets, but the measurements of the excitation potentials were not sufficiently precise to distinguish between the two critical potentials.

Excitation potentials corresponding to the other lines in the principal series have not yet been observed. We should expect the ionization potential to agree with that calculated from the limit of the principal series because the successive lines are supposed to correspond to states in which the outer electron is farther and farther from the nucleus, so that the limit should correspond to the case when the electron is entirely removed from the atom.

The following table contains the observed and calculated critical potentials of helium in volts.

Calculated		Observed
19·73	...	19·73
20·56	. .	20·53
21·12	.	21·2
22·98		22·9
24·48	. .	24·5

The calculated values are got from the wave numbers of spectral lines believed to be emitted when the helium atom reverts to its normal state, except the last number, which is got from the limit of the corresponding series, since 24·5 volts is the ionization potential.

In the case of mercury, eighteen critical potentials have been observed ranging from 4·68 volts to the ionization potential 10·39 volts. Most of these agree with values calculated from the wave numbers of spectral lines believed to be emitted by mercury atoms reverting to the normal state. Similar results have been obtained with many other elements.

6. Smyth's Application of Positive Ray Analysis.

Very interesting and important results have been obtained by Smyth by means of a combination of Lenard's method of measuring critical potentials and positive ray analysis. In this way it is possible to determine the nature of the positive ions produced in gases by electron collisions.

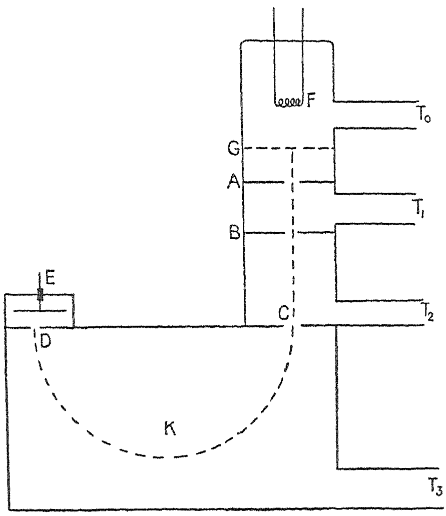


Fig. 6

The most essential parts of Smyth's apparatus are shown diagrammatically in fig. 6. Electrons from a filament F are accelerated to a grid G and then stopped by a retarding potential between the plate A and the grid. Positive ions produced by electron collisions between G and A are accelerated towards A and some of them pass through small holes in the diaphragms A, B, and C into a box K where they are deflected by a uniform magnetic field perpendicular to the plane of the paper. In this field they describe circular paths and by adjusting the field strength may be made to enter a slit at D, where they fall on an insulated electrode E. The charge received by E is measured with a quadrant electrometer. The box K and the spaces between the diaphragms A, B, and C are kept highly exhausted by means of powerful pumps connected to the tubes T_1, T_2, and T_3, so that the positive ions are not stopped by collisions with gas molecules. A slow stream of the gas to be investigated is passed into the apparatus through the tube T_0, making it possible to have sufficient gas pressure in the space between the grid and plate to obtain enough ions by collisions although the rest of the apparatus is kept highly exhausted.

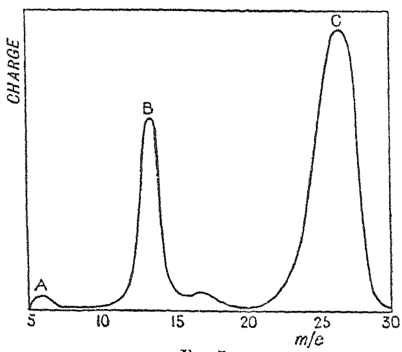


Fig. 7

The accelerating potential V_a between the filament and grid is kept constant, as also the retarding potential between G and A. The potential between A and B which accelerates the positive ions produced between G and A is varied and the magnetic field kept constant. The values of m/e for the positive ions can be easily calculated from the radius of the path which they describe, the potential accelerating them, and the strength of the magnetic field.

In fig. 7 the way in which the charge received by the electrode E varies with the value of m/e for the positive ions reaching E is shown. In this case the gas used was nitrogen and the accelerating potential V_a was 100 volts. The value

of m/e is expressed in terms of m/e for protons, so that it is equal to the atomic weight in the case of ions carrying a single electronic charge

The curve shows three maxima at m/e equal to 7, 14, and 28. These must clearly be due to positive ions consisting of atoms of nitrogen with two electronic charges, atoms with one electronic charge, and molecules N_2 with one electronic charge. The relative amounts of these three sorts of positive ions given by nitrogen was determined for a series of values of the accelerating potential V_a. It was found that when V_a is less than 25 volts only molecules N_2^+ with one electronic charge are obtained. Above 25 volts small amounts of N^+ and N^{++} appear This shows that when nitrogen gas is ionized by collisions with electrons practically all the positive ions produced are N_2^+. The nitrogen molecule is evidently a very stable system not easily dissociated into two atoms. This result explains the chemical inertness of nitrogen

Very similar results were obtained with hydrogen H_2^+ ions were readily formed but H^+ was obtained more easily than N^+.

The critical potentials of atoms corresponding to the production of characteristic X-rays are discussed in the chapter on X-rays.

Reference

"Critical Potentials." K. T. Compton and F. L. Mohler. *Bulletin of the National Research Council*, Washington, D.C.

CHAPTER VII

X-Rays and γ-Rays

1. Nature of X-rays.

X-rays or Rontgen rays were discovered by Röntgen in 1895. He was using a Crookes tube in which the cathode rays struck the glass walls of the tube, and he noticed that a piece of paper coated with barium-platinocyanide, which happened to be lying near, fluoresced when the tube was working On putting objects between the tube and the paper, shadows were obtained in the fluorescent light, showing that the tube was emitting some kind of radiation. It was found that this radiation was remarkably penetrating. It passed readily through black paper and thin sheets of aluminium which were quite opaque to ordinary light. It was found that the new radiation could not be refracted or reflected but travelled along straight lines through objects of any shape, and it was not deflected by either magnetic or electric fields.

J. J. Thomson discovered that the X-rays produce electrical conductivity in gases and other insulators when passed through them.

It was found that the X-rays are emitted where the cathode rays strike any solid object In 1881 J. J. Thomson pointed out that if cathode rays are rapidly moving charged particles, then, according to Maxwell's electromagnetic theory, they should produce electromagnetic radiation when suddenly stopped. He suggested that the green fluorescence of the glass walls of Crookes tubes produced by the cathode rays may be produced by this radiation emitted when the rays strike the glass. Soon after the discovery of X-rays Stokes and Wiechert suggested that the sudden stopping of the cathode rays when they strike a solid body causes the emission of electromagnetic pulses or waves of very short wave-length which are the X-rays. It is believed now that the cathode rays penetrate the atoms of the solid body and are not stopped suddenly. The cathode rays ionize and excite the atoms by collisions and the X-rays are emitted when the atoms revert to their normal state. X-rays are simply light or electromagnetic waves of very short wave-length. Barkla in 1905 showed that polarized X-rays can be obtained, and it is now known that X-rays can be

diffracted, reflected, and refracted, and so have all the properties of light on which the wave theory of light is based.

2. The Coolidge Tube.

The X-rays from ordinary Crookes tubes are very feeble, and powerful X-rays are now obtained from specially designed tubes. The most satisfactory X-ray tubes are the Coolidge tubes, in which the electrons emitted by a hot filament are accelerated towards a metal anode usually made of tungsten or molybdenum

A Coolidge tube is shown diagrammatically in fig. 1.

The cathode C consists of a small flat spiral of tungsten wire surrounded by a tungsten cylinder. The spiral is heated by a current

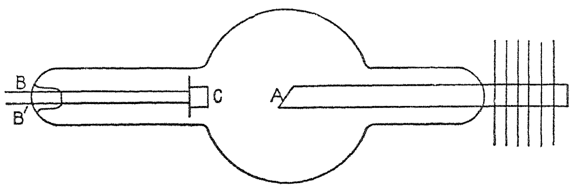


Fig. 1

from a battery through the wires B and B′ sealed through the glass. The anode A consists of a plug of tungsten or molybdenum set in the end of a copper rod which is cut off at 45° as shown. The other end of the copper rod, outside the bulb, is cooled by means of several copper discs which expose a large surface to the air. The bulb is very completely exhausted, so that no electrical discharge can be passed through it when the filament is cold.

A difference of potential of about 50,000 to 100,000 volts is maintained between the anode and cathode, and the filament heated so that it emits electrons The current through the tube can be measured with a milliammeter and may be up to 50 or 100 milliamperes. The electrons emitted by C are concentrated on to a small area on the end of A by the electrostatic repulsion of the charge on the end of the cylinder surrounding the filament. The X-rays are emitted by this small area about equally in all directions. They are absorbed of course by the anode when they go into it. The anode of a tube taking. say, 50 milliamperes at 100,000 volts or 5 kilowatts rapidly gets hot, for nearly all the energy of the electrons is converted into heat when the electrons are absorbed by the anode. Less than 1 per cent of the energy goes into the X-rays.

3. Scattering. Diffraction. Characteristic X-rays.

When X-rays are passed through matter they are partly scattered in all directions, while at the same time the matter may be caused to

emit secondary X-rays different in quality from the incident rays. Barkla discovered that the secondary X-rays emitted by different elements have properties characteristic of the particular element emitting them which vary in a regular way with the atomic weight of the element. X-rays also cause matter to emit electrons, an effect analogous to the emission of electrons due to ultra-violet light.

In 1912 Laue discovered that X-rays can be diffracted by crystalline substances, which act on the rays in the same sort of way as a diffraction grating acts on light. By means of this effect it is possible to determine the wave-lengths of X-rays, and it is found that the X-rays emitted by different elements have definite wave-lengths characteristic of the element, just like the wave-lengths of the spectral lines in the optical spectra of the elements. Laue's important discovery placed the study of X-rays on a new quantitative basis and will now be considered in detail.

X-RAYS AND CRYSTAL STRUCTURE

4. Crystallography. Law of Rational Indices.

Crystals are homogeneous solid bodies bounded by plane surfaces which are called the faces of the crystal. The faces are arranged in sets, all those in a set being perpendicular to the same plane; these sets are called zones. The faces in a zone intersect in parallel lines. The faces occur in parallel pairs on opposite sides of the crystal.

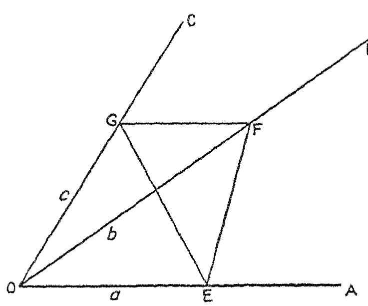


Fig. 2

The geometrical character of the crystals of any substance is determined by the angles between their faces, which are invariable, and is independent of the size of the particular crystal selected. Crystals grow by the deposition of uniform layers on their faces, so that they get larger without the angles between the faces changing.

The different faces of a crystal can be specified by the intercepts cut off from three axes by the faces. As axes, lines parallel to three edges of the crystal which do not lie in one plane are used.

In fig. 2, let OA, OB, and OC be the three axes selected, and let E, F, G be the points at which a face of the crystal intersects them. Let OE $= a$, OF $= b$, and OG $= c$. The ratios $a : b : c$ will always be the same for the corresponding axes and face on any crystal of the same substance, whatever the size of the crystal. Any other plane parallel to the face EFG will cut off lengths from the axes in the same ratios as $a : b : c$ and could equally well be taken to be the face of the crystal.

Now let a', b', and c' be the intercepts for any other face of the crystal; then it is found that there is always a simple numerical relation between a', b', c' and a, b, c. The ratios $a' : b' : c'$ are equal to the ratios $a/k : b/h : c/l$, where k, h, and l are integers which are usually quite small and are called the indices of the face they specify. Thus

$$\frac{a/k}{b/h} = \frac{a'}{b'}, \text{ so that } \frac{h}{k} = \frac{b}{a}\frac{a'}{b'},$$

and in the same way

$$\frac{l}{h}=\frac{c}{b}\frac{b'}{c'} \text{ and } \frac{k}{l}=\frac{a}{c}\frac{c'}{a'}.$$

If $k=h=l=1$, we have $a' : b' : c' = a : b : c$, and so have the original face. If $k=h=1$ and $l=2$ we have $a' : b' : c' = a : b : c/2$; and if $k=1, h=2$, and $l=3$ we have $a' : b' : c' = a : b/2 : c/3$, and so on. Thus any face of the crystal can be specified by giving the values of k, h, and l for that face. The values of k, h, l, of course, depend on the particular edges chosen as axes and on the particular face selected as the $k=h=l=1$ face.

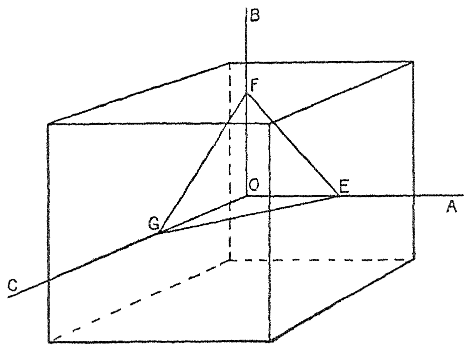


Fig 3

If a face is parallel to one of the axes, as frequently happens, then its intercept on that axis is infinitely long, so that the integer becomes zero. Thus if $k=0$, $h=1$, $l=1$ so that $a'=a/0=\infty$, $b'=b$, $c'=c$, the face is parallel to the axis OA. The indices k, h, and l can be positive or negative integers or zero. Any face of a crystal can therefore be specified by giving the values of h, k, and l for that face. The numbers are usually put inside a bracket, thus (112) denotes the face with intercepts a, b, and $c/2$. Negative values are indicated by a minus sign above the number thus $(11\bar{2})$.

By properly selecting the axes and reference face (111) it is usually possible to represent all principal faces by quite small values of h, k, and l.

As an example, consider a cube (fig. 3) and let the origin be at its centre and the axes parallel to three of its edges which meet at one corner. In this case, if $a=b=c$, the six sides of the cube are (100), $(\bar{1}00)$, (010), $(0\bar{1}0)$, (001), $(00\bar{1})$, since each side is perpendicular to one axis and parallel to the other two. The reference face EFG passes through the three points where the positive directions of the axes cut the surface of the cube. It is not one of the faces of the cube, but is a possible face for a crystal having faces parallel to the sides of the cube. Eight possible faces are (111), $(1\bar{1}1)$, $(11\bar{1})$, $(1\bar{1}1)$, $(\bar{1}11)$, $(1\bar{1}\bar{1})$, $(\bar{1}11)$, $(\bar{1}11)$. These eight faces are shown in fig. 4. They form an octahedron.

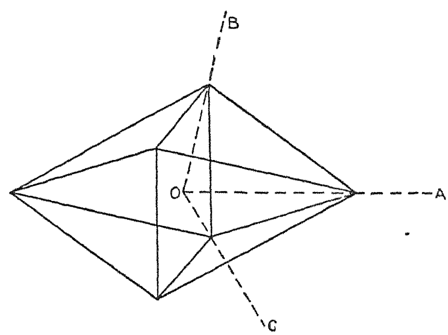


Fig 4

The fact that the indices are positive or negative integers or zero for any face of a crystal is known as the law of rational indices. This law, which was first discovered empirically, can be explained in a simple way on the atomic theory. We suppose that crystals consist of a regular arrangement of atoms, the same throughout the volume of the crystal. The atoms are arranged in groups or sets, all groups being similar and similarly orientated. In a crystal of a compound

body each group will usually consist of the atoms forming one molecule of the compound. In crystals of an element each group may consist of a single atom. The position of a group is fixed by the position of any definite point in it, for example, the centre of one of the atoms in it. We shall for the present speak of each group as being located at the point fixing its position. The groups must be arranged in the crystal in such a way that every group is related to its neighbouring groups in the same way. Thus if A_3 and A_2 are two groups in a crystal (fig. 5) then there must also be groups at A_1, A_2, A_5, A_6 on the line joining A_3 and A_2, and the groups must be equally spaced along this line. If B_1 is another group near A_1, then there must also be groups at B_2, B_3, B_4, &c., in such positions that $A_2B_2 = A_1B_1$, $A_1A_2 = B_1B_2$, &c. In this way it is easy to see that all the groups in a crystal must be arranged at the intersections of three sets of equally spaced parallel planes. Such an arrangement is called a space lattice.

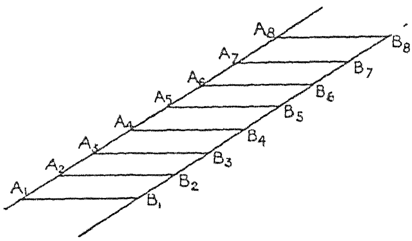


Fig. 5

Three sets of equally spaced parallel planes divide space up into equal parallelepipeds as shown in fig. 6, and there is one group of atoms at each point where three planes intersect. The number of groups of atoms is equal to the number of the parallelepipeds.

If a plane is drawn through any three groups of atoms in the space lattice, the plane will contain groups of atoms arranged at the intersections of two sets of parallel lines, and it is supposed that such a plane is a possible face of the crystal. The faces usually observed are those determined by planes which pass through three groups of atoms which are near together in the space lattice. If three sides of one of the parallelepipeds, meeting at a point, are taken as the axes then any plane drawn through three groups will have intercepts on these axes which are multiples of the lengths of the sides of the parallelepipeds. Let these lengths be a_1, a_2, a_3 so that the intercepts are n_1a_1, n_2a_2, n_3a_3, where n_1, n_2, n_3 are integers. For another plane through three other groups let the intercepts be $n_1'a_1$, $n_2'a_2$, $n_3'a_3$. If we put $n_1a_1 : n_2a_2 : n_3a_3 = n_1'a_1/h : n_2'a_2/k : n_3'a_3/l$, we have $h/k = n_2'n_1/n_1'n_2$, $l/h = n_2n_3'/n_3n_2'$; but the n's are all integers and therefore the ratios h/k and l/h can be expressed as the ratios of integers. Thus we see that the law of rational indices follows from the theory that crystals consist of groups of atoms arranged in space lattices and that the faces are planes drawn through the groups. The axes and reference face (111) are chosen for any particular crystal so that the other faces are given by the simplest possible sets of indices. The principal faces of a crystal can usually be specified without using integers greater than 2. It should be observed that there are many different sets of equally spaced parallel planes the intersections of which form any particular space lattice.

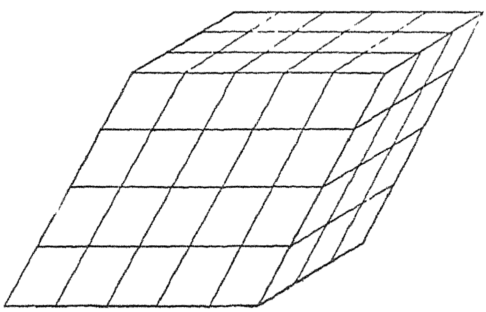


Fig. 6

5. Diffraction Patterns.

If we draw planes through every group in a crystal, all parallel to one of the faces, we get a set of equally spaced parallel planes. Since all the groups of atoms are supposed to be similar and similarly orientated it follows that the arrangement of atoms is repeated at regular equal intervals measured from the face along any normal to it. When X-rays are passed through matter they are partly scattered in all directions. It is supposed that this scattering is produced by the electrons in the atoms. If the X-rays are a train of waves of definite wave-length λ, each electron is supposed to be made to vibrate with the frequency c/λ of the rays by the electric field in the rays. The electron therefore emits radiation of the same wave-length as the rays falling on it.

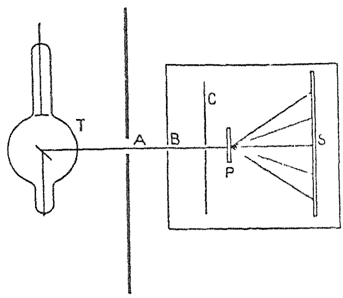


Fig 7

If a narrow beam of X-rays is passed through a crystal, then in certain directions the scattered rays from the regularly arranged atoms are all in phase and together form a comparatively intense scattered beam. The apparatus used in Laue's original experiment is shown in fig. 7. The rays from a tube T are passed through a series of small holes in lead screens A, B, C, and the narrow pencil of rays falls on a crystalline plate P about 1 mm. thick A photographic plate S is put up a few centimetres behind the crystal. After an exposure of several hours to the X-rays, on developing the plate a symmetrical pattern of spots is found on it. These spots are arranged round a central spot formed by the rays which pass straight through the crystal. Some of the diffracted rays may make angles of 40° or more with the incident pencil. Fig. 8 shows reproductions of such diffraction patterns. (See Plate.)

6. W. L. Bragg's Theory of Diffraction Patterns.

The theory of these diffraction patterns was worked out by Laue and was found to agree exactly with the facts Laue's theory was complicated, and a greatly simplified form of the theory was proposed by W. L. Bragg. Bragg's theory gives the same results as Laue's and will be considered here.

Consider a narrow beam of X-rays from a source S (fig. 9) falling on a plane EF, and suppose that single atoms are distributed over this plane. Each atom scatters a minute fraction of the rays falling on it in all directions. Suppose that the scattered waves from two atoms at A and at A' which are near together arrive at a point P at the same

time so that the waves reinforce each other at P. For this to happen we must have

$$SA + AP = SA' + A'P.$$

Now if $SA + AP =$ constant, A must be on the surface of an ellipsoid of revolution about SP with foci at S and P. If the plane EF is a tangent plane to this ellipsoid at A, the scattered waves from all the atoms on the plane EF which are very near A will arrive at P together, so that there will be regular reflection of a very small fraction of the X-rays along the path SAP. If EF is a tangent plane to the ellipsoid, then SA, AP, and the normal AN to EF at A lie in the same plane, and the angle SAN is equal to the angle NAP.

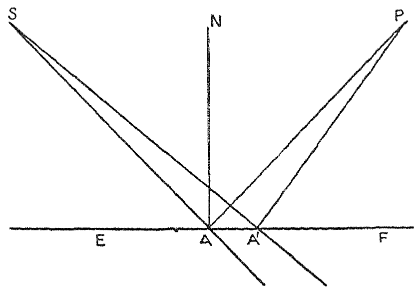


Fig 9

Now suppose that the plane EF is on the face of a crystal. As we have seen, the atoms in the crystal are regularly arranged in groups about equally spaced parallel planes of which the face is one. Let the distances between these equally spaced plane layers of atoms be d. Then each layer will regularly reflect a small fraction of the rays falling on it. If the rays reflected from all the layers are in the same phase and so reinforce each other a strong reflected beam will be obtained. In fig. 10 let S be the source of the rays and A_1, A_2, A_3, &c.,

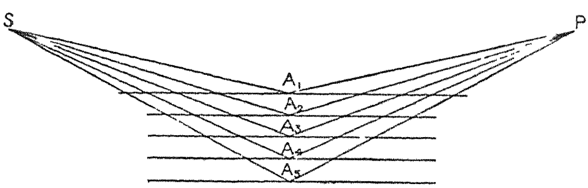


Fig 10

the equally spaced layers of atoms. Let the rays from S be reflected at A_1, A_2, A_3, &c., to P, and suppose that the reflected wave trains from each plane arrive at P in the same phase and so reinforce each other at P. This requires that

$$(SA_2 + A_2P) - (SA_1 + A_1P) = n\lambda,$$
$$(SA_3 + A_3P) - (SA_2 + A_2P) = n\lambda,$$
$$(SA_4 + A_4P) - (SA_3 + A_3P) = n\lambda,$$
$$\cdots\cdots\cdots\cdots\cdots$$

where n is a positive integer; that is, the path differences for the successive planes must be a multiple of the wave-length λ. The distance d between the reflecting planes is very small compared with SA and AP so that the rays near A are practically parallel. The path differences are therefore easily calculated. In fig. 11 let p and p' be two successive reflecting planes so that $AA' = d$, and let the angles ANA' and AMA' be right angles. The path difference is $NA' + A'M$, which is equal to $2d \sin\theta$ where θ is the angle between the incident or the reflected rays and the reflecting planes. The condition for strong reflection from the crystal face is therefore

$$2d \sin\theta = n\lambda,$$

where $n = 1, 2, 3 \ldots.$

Such reflection of X-rays at a crystal face, of course, is not a reflection at the surface of the crystal but is reflection by a very large number of the equally spaced layers of atoms. The distances between the layers are of the order of magnitude 10^{-8} cm., so that there are a million layers in a thickness of the order of $\frac{1}{10}$ mm. The angle θ is usually called the glancing angle, and d is often called the crystal grating space. There is no appreciable reflection at angles differing even very slightly from the values given by $2d \sin\theta = n\lambda$. For if the path difference is $\lambda\left(n + \frac{1}{2m}\right)$, instead of $n\lambda$, where m is a large whole number, then there is a path difference $\lambda(nm + \frac{1}{2})$ between the rays reflected from any plane and the mth plane below it. The number nm is an integer, so that the two wave trains interfere and destroy each other. The reflected wave trains therefore interfere in pairs and there is no appreciable reflection, unless m is comparable with the total number of reflecting planes, which is very large. When a beam of X-rays is passed through a crystal as in Laue's experiment, the rays may be reflected from all the sets of planes into which the groups of atoms can be supposed arranged. If the X-rays are not homogeneous but have a continuous spectrum so that all wave-lengths between certain limits are present, then each set of planes reflects the rays of wave-length λ given by

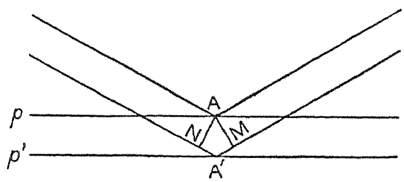


Fig 11

$$2d \sin\theta = n\lambda,$$

where θ is the glancing angle between the set of planes and the incident beam. In this way the diffraction patterns obtained can be explained.

The intensity of the reflection from any set of layers increases with the number of atoms per square centimetre in the layers, so that, other

things being equal, sets for which d is small reflect less than those for which d is large. The more intense spots in the patterns are those from sets of planes which can be drawn through three groups of atoms which are close together in the crystal.

7. W. H. Bragg's X-ray Spectrometer.

The reflection of X-rays from the faces of crystals enables the wave-lengths of the rays to be found and also the arrangement of the atoms in the crystal This method of studying X-rays and crystals is due to W. H. Bragg The instrument used for measuring the glancing angles is called an X-ray spectrometer.

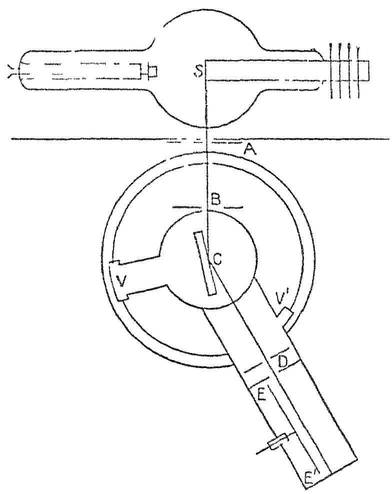


Fig 12

Fig. 12 shows the essential parts of an X-ray spectrometer used by W. H. Bragg. The X-rays from the anti-cathode S of a Coolidge tube pass through two slits A and B and fall on the face of a crystal at C. The crystal is mounted on a table which can be rotated about an axis in the crystal face and the angles through which it is turned can be read off on a scale by the vernier V. The reflected rays pass through a slit at D and then through a thin aluminium window into an ionization chamber EE′. The slit D and chamber can be turned about the same axis as the crystal and the angles read off by the vernier V′.

The ionization chamber is a cylindrical lead box about 15 cm. long and 5 cm. in diameter. The narrow beam of rays passes along the centre of the chamber and ionizes the gas in it. The chamber contains an insulated electrode, which is connected to some form of sensitive electrometer. C. T. R. Wilson's tilted gold leaf electroscope is frequently used. The chamber is charged to about 100 volts by connecting it to a battery, and the insulated electrode is kept near zero potential. The conductivity produced in the chamber by the X-rays can be measured with the electrometer. The chamber is usually filled with ethyl bromide vapour or some other gas which absorbs the rays strongly.

To study the reflection of the X-rays by the crystal, the chamber slit is set so as to receive the reflected rays. The angle between the chamber slit and the incident rays must be twice the angle between the

crystal face and the incident rays. If the ionization is measured for a series of glancing angles a curve may be drawn showing the variation of ionization with the angle. Fig 13 shows such a curve obtained by Bragg with the X-rays from a rhodium anticathode and a diamond crystal The ionization shows two maxima at about 36·5° and 36·9° which are due to two lines in the X-ray spectrum of rhodium. The small amount of ionization obtained at other angles is due to the continuous part of the X-ray spectrum.

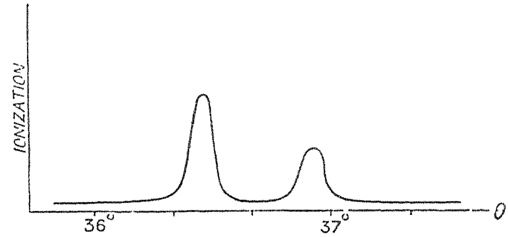


Fig 13

Instead of using an ionization chamber the reflected rays may be received on a photographic plate and the angles found by measuring the distances between the images on the plate. The ionization method has the advantage that it gives the relative intensities of the lines accurately. Any particular line is only reflected when the crystal is set so that the glancing angle satisfies the equation

$$n\lambda = 2d \sin\theta.$$

To photograph all the lines which can be reflected from a crystal face it is therefore necessary to slowly rotate the crystal backwards and forwards during the exposure of the plate. In this way a photograph of all the lines reflected in the range of angles covered can be obtained. Each line can be reflected at all the angles given by $n\lambda = 2d \sin\theta$, with $n = 1, 2, 3, 4$, &c. $n\lambda$ of course must be less than $2d$, since $\sin\theta$ cannot be greater than 1. The lines are usually observed on both sides of the incident beam, and half the angle between the two positions is taken as the value of θ so as to eliminate zero errors.

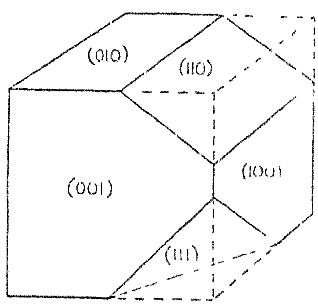


Fig 14

8. Example of Use of X-ray Spectrometer.

As a simple example of the use of the X-ray spectrometer to determine the structure of crystals and the wave-lengths of X-rays we will consider the reflection of the X-rays from a palladium anticathode by crystals of sodium and potassium chlorides A full account of the subject is given in *X-rays and Crystal Structure*, by W. H. and W. L Bragg

Sodium chloride and potassium chloride can be obtained in cubical crystals the faces of which are specified by the indices (100), (010), (001) Besides these faces the faces (110) and (111) often appear. Fig. 14 shows these faces.

The (100) face of KCl gives strong lines at 5° 23′, 10° 49′, 16° 20′. The sines of these angles are as 1 : 2 : 3, so that we conclude that they correspond to the same wave-length λ with $n = 1$, 2, and 3.

With the (110) and (111) faces the angles for $n = 1$ are 7° 37′ and 9° 23′. Now $\sin 5° 23' : \sin 7° 37' : \sin 9° 23' = 1 : \sqrt{2} : \sqrt{3}$ almost exactly. It appears therefore that the grating spaces for the faces (100), (110), and (111) are as $1 : \frac{1}{\sqrt{2}} : \frac{1}{\sqrt{3}}$. This is what we should obtain if the space lattice of the crystal consisted of equal cubes with an atom or group of atoms at each corner. The grating space for the (100) face would then be equal to the edge of a cube, for the (110) face to half the diagonal of a face of the cube and for the (111) face to one-third of the diagonal of the cube. These lengths are as $1 : \frac{1}{\sqrt{2}} : \frac{1}{\sqrt{3}}$ in agreement with the results obtained. We conclude that the crystals of KCl behave like a simple cubical space lattice.

The glancing angle for the same X-ray line from the (100) face of sodium chloride is 6° 0′. The grating space for this face is therefore smaller than for potassium chloride in the ratio $\sin 5° 23' : \sin 6° 0' = 1 : 1.115$. If the NaCl and KCl crystals have the same structure, which we should expect to be the case, then the grating spaces should be proportional to the cube roots of the molecular volumes.

The molecular volume of KCl is equal to the molecular weight $39{\cdot}10 + 35{\cdot}46 = 74{\cdot}56$ divided by its density 1·99 or 37·5 and that of NaCl is $\frac{23{\cdot}00 + 35{\cdot}46}{2{\cdot}163} = 27{\cdot}02$ and $\left(\frac{37{\cdot}5}{27{\cdot}0}\right)^{1/3} = 1{\cdot}115$, which agrees with the ratio of the grating spaces exactly.

The glancing angle for the (110) face of NaCl agrees with that for KCl in the same way, but for the (111) face of NaCl the glancing angle is just one-half that to be expected from the angle for KCl. In the case of NaCl the grating spaces for the faces (100), (110), and (111) are as $1 : \frac{1}{\sqrt{2}} : \frac{2}{\sqrt{3}}$, instead of $1 : \frac{1}{\sqrt{2}} : \frac{1}{\sqrt{3}}$ as with KCl. The numbers of electrons in K, Cl, and Na atoms are 19, 17, and 11 respectively, so that K and Cl atoms should be nearly equally efficient as scatterers of X-rays but Na atoms little more than half as efficient. Thus in a KCl crystal all the atoms are nearly equal as regards X-ray scattering, so that it makes little difference whether a K atom or a Cl atom is at any particular point. This suggests that in the simple cubical lattice of a KCl crystal the group is either one K atom or one Cl atom, and that in a NaCl crystal the arrangement of the atom is the same as in a KCl crystal, but that the effect on X-rays is not the same because of the difference between the number of electrons in Na and Cl atoms.

Since the grating spaces for the (100) and (110) faces correspond we conclude that the planes parallel to these faces contain equal numbers of metal and chlorine atoms in each case, whereas in the case of the (111) faces the planes contain all metal atoms and all chlorine atoms alternately. In the case of KCl the (111) grating space is the distance from a plane containing only K atoms to the next plane containing only Cl atoms, since these planes are practically equal as scatterers, but in the case of NaCl crystals the (111) grating space is the distance from a plane containing only Na atoms to the next plane containing only Na atoms and so is twice that corresponding to the (111) face in KCl. The arrangement of the atoms is therefore as shown in fig. 15. The black circles represent one sort of atom and the white circles the other sort. The atoms with equal numbers are all in the same (111) plane. Each sort of atom is arranged in a face-centred cubical lattice, that is, at the corners of cubes and in the middle of each face.

The relative intensities of the reflected rays for different values of n in the equation $n\lambda = 2d \sin\theta$, or the relative intensities of the spectra of different orders, also give information about the arrangement of the atoms. For example, it is found that with the (111) face of sodium chloride the lines for which $n = 2$ are stronger than those for which $n = 1$. The grating space in this case is the distance from a layer of Cl atoms to the next layer of Cl atoms. The rays reflected from the layers of Na atoms in between, when $n = 1$, are half a wave-length behind the waves from the next layer of Cl atoms and so diminish the intensity of the reflected beam but, when $n = 2$, the rays from the Na atoms are in phase with the rays from the Cl atoms and so a strong reflected beam is obtained. In the case of potassium chloride the rays from the layers of K atoms are practically equal to the rays from the layers of Cl atoms, so that if d is regarded as the distance from a Cl layer to the next Cl layer, then when $n = 1$ there is practically no reflected beam, so that d is apparently the distance from a Cl layer to a K layer, as we have seen. These considerations serve to support the assumed arrangement of the atoms in these crystals.

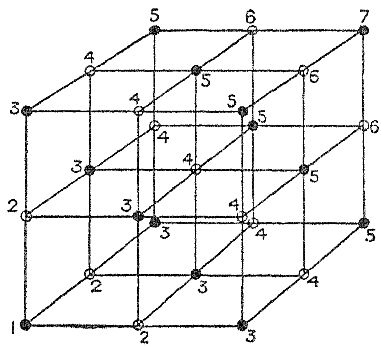


Fig 15

When the arrangement of the atoms in a crystal has been decided, the absolute values of the grating spaces can be calculated from the known masses of the atoms and the density of the crystal. The mass of a NaCl molecule is $\frac{58{\cdot}46}{1{\cdot}0077} \times 1{\cdot}663 \times 10^{-24}$ grams, since $1{\cdot}663 \times 10^{-24}$ is the mass of one atom of hydrogen of atomic weight 1·0077. This gives $9{\cdot}797 \times 10^{-23}$ grams. The volume occupied by one atom of either sodium or chlorine in the crystal is therefore $\frac{9{\cdot}797 \times 10^{-23}}{2 \times 2{\cdot}163} = 22{\cdot}65 \times 10^{-24}$ c. c. The grating space for the (100) face is equal to the edge of the elementary cubes, of which there is one to each atom, so that

$$d_{(100)} = (22{\cdot}65 \times 10^{-24})^{1/3} = 2{\cdot}83 \times 10^{-8} \text{ cm.}$$

The wave-length of the X-ray spectral line for which $\theta = 6^\circ 0'$ when $n = 1$ is therefore

$$\lambda = 2 \times 2{\cdot}83 \times 10^{-8} \times \sin(6^\circ 0') = 0{\cdot}592 \times 10^{-8} \text{ cm.}$$

When the grating space for a face of a crystal has been found, the crystal can be used to determine the wave-lengths of X-ray spectral lines. The crystals usually employed are calcite ($d = 3{\cdot}028 \times 10^{-8}$ cm.) or rock salt ($d = 2{\cdot}814 \times 10^{-8}$ cm.).

9. Wave-lengths of Characteristic X-rays of the Elements. Moseley's Work.

The X-rays of any element are obtained by using it or one of its compounds as the anti-cathode of an X-ray tube such as the Coolidge tube. The elements can also be made to emit their characteristic X-rays by exposing them to X-rays of sufficiently short wave-length.

The wave-lengths of the lines in the X-ray spectra of many of the

elements were first measured by Moseley, who discovered the very important fact that the frequencies vary in a regular way with the atomic number.

The X-ray spectral lines of any element can be arranged in series or groups of lines. The series having the highest frequencies is called the K series, that with the next highest frequencies the L series, and so on. The K series usually contains four strong lines known as K_{α_2}, K_{α_1}, K_β, and K_γ. The following table gives the wave-lengths of these K series lines in Ångstrom units ($=10^{-8}$ cm.) for several elements, and the atomic numbers of these elements. The atomic number of an element, except in one or two cases, is the number giving the position of the element in a list of the elements arranged in the order of increasing atomic weights. Thus the atomic numbers of hydrogen, helium, and lithium are 1, 2, and 3 respectively.

Element.	Atomic Number	Wave-lengths of K Series			
		α_2	α_1	β	γ
Sodium	11	–	11·8836	11·591	–
Potassium .	19	3·738	3·7339	3·4464	–
Iron . .	26	1·932	1·9324	1·7540	1·736
Bromine .	35	1·040	1·035	0·929	0·914
Rhodium .	45	0·6164	0·6121	0·5453	0·5342
Caesium . .	55	0·402	0·398	0·352	– –
Tungsten .	74	0·2131	0·2086	0·1842	0·17901

Moseley found that if the square roots of the frequencies of the K_α lines are plotted against the atomic numbers then a curve is obtained which is very nearly straight. Approximately, if ν_{K_α} is the frequency,

$$\nu_{K_\alpha} = \tfrac{3}{4}R(N-a)^2,$$

where N is the atomic number, a is a constant, and R is a constant which is equal to the Rydberg constant of the formulæ for optical spectral series. For example, the frequencies of the series in the spectrum of atomic hydrogen are given by $\nu = R\left(\frac{1}{n^2}-\frac{1}{m^2}\right)$, where n and m are integers. For the strongest line of the L series, the L_α line, Moseley found

$$\nu_{L_\alpha} = \tfrac{5}{36}R(N-b)^2.$$

Now $$\frac{3}{4}=\frac{1}{1^2}-\frac{1}{2^2} \text{ and } \frac{5}{36}=\frac{1}{2^2}-\frac{1}{3^2}.$$

10. **Quantum Theory of Spectra. Energy Levels. K, L, M, and N Series. Quantum Numbers.**

According to Bohr's quantum theory, the frequencies of the spectral lines of an atom consisting of a nucleus with a positive charge Ne and a single electron are given (Chap. V, section 11) by

$$\nu = RN^2\left(\frac{1}{n^2} - \frac{1}{m^2}\right).$$

Moseley's formulæ for the K_a and L_a lines are clearly closely related to this last equation, and it was immediately seen that Moseley's results indicate that the atomic numbers of the elements are equal to the nuclear charges expressed in terms of the protonic charge e. This has been confirmed by all subsequent work, and in particular by researches on the scattering of α-rays.

According to the quantum theory of spectra the frequencies of the lines are given by

$$h\nu = W_m - W_n,$$

where h is Planck's constant, and W_m and W_n denote the energies of the atom in two of its possible states. The radiation is emitted when the atom changes from a state with energy W_m to one with less energy W_n, the change being supposed to consist of an electron dropping from an outer possible orbit to an inner vacant orbit. In the case of an atom having only one electron outside the nucleus we have

$$\frac{W_n}{h} = A - \frac{RN^2}{n^2},$$

so that

$$\frac{W_m - W_n}{h} = RN^2\left(\frac{1}{n^2} - \frac{1}{m^2}\right) = \nu.$$

For atoms with more than one electron outside the nucleus no theoretical expression for the energies of the possible states or energy levels of the electrons has been worked out, but since every frequency emitted gives the difference between two energy levels it is possible to deduce the energy levels of an atom from the frequencies of the lines which it emits.

The principal lines in the X-ray spectra of any element can be regarded as due to transitions of electrons between energy levels which are called the K, L, M, and N levels. The K series is due to transitions from the L, M, and N levels to the K level, the L series to transitions from the M and N levels to the L level, and the M series to transitions from the N level to the M level. Each of these levels except the K level is really not a single level but a group of levels which

do not differ much from each other. The number n in the equation

$$\frac{W_n}{h} = A - \frac{RN^2}{n^2}$$

is called the *quantum number* fixing the state of the atom, and the state is referred to as an n-quantum state, or an n-quantum energy level, of the electron. This terminology is carried over to the more complicated atoms, and the K, L, M, and N states or energy levels are referred to as one, two, three, and four quantum states or energy levels respectively.

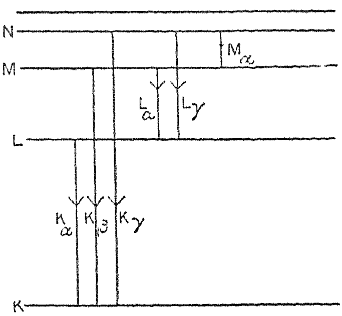


Fig 16

The energy levels and the transitions between them are shown diagrammatically in fig 16. The horizontal lines represent the energy levels, and the vertical distances between them the energy differences such as $W_L - W_K$. The vertical distance between the K level and the line at the top is supposed to represent the energy required to remove an electron from the K level right out of the atom. The X-ray spectral lines are represented by the vertical lines, which show a transition from the level at the top of the line to that at the bottom. We see at once that

$$\nu_{K_\alpha} + \nu_{L_\alpha} = \nu_{K_\beta}, \quad \nu_{K_\alpha} + \nu_{L_\gamma} = \nu_{K_\gamma}, \quad \nu_{L_\alpha} + \nu_{M_\alpha} = \nu_{L_\gamma},$$

with other similar relations. It is found that these relations between the frequencies are accurately true.

11. X-ray Absorption Spectra. Energy Levels.

The absorption spectrum of an element for X-rays can be observed with the X-ray spectrometer by passing the X-rays from a Coolidge tube through a plate of the substance, and examining the spectrum of the transmitted rays either photographically or with the ionization chamber. It is found that there are certain critical wave-lengths at which the absorption suddenly increases as the wave-length diminishes. A critical absorption wave-length characteristic of an element is a wave-length such that the element absorbs X-rays of shorter wave-length than the critical value more than it does wave-lengths longer than the critical value. If the element is present in a gas through which the X-rays are passed, the ionization of the gas is greater for X-rays with wave-lengths shorter than the critical value than for longer wave-lengths. The increased absorption is accompanied by increased ionization. According to the quantum theory, absorption

occurs when an electron in the atom is moved from its orbit or energy level to the free state outside the atom. In a normal atom all the energy levels are occupied by electrons, so that it is not possible for the radiation to move an electron from one level to another. If E_n denotes the energy required to move an electron from an n quantum level to a position outside the atom, then the incident radiation can only move the electron if its frequency ν is at least as great as that given by $h\nu = E_n$. When an electron drops back from just outside into the vacant level the atom emits a quantum of radiation $h\nu = E_n$. The critical absorption frequencies of an element therefore give the values of E_n for the element. The K series, for example, is due to electrons falling from the L, M, N, &c., levels into the K level. The frequency of the line in the K series of highest frequency should therefore be practically identical with the critical absorption frequency corresponding to the K level, for the highest energy level must nearly correspond to a point just outside the atom. It is found in fact that the K critical absorption frequencies are only a fraction of one per cent greater than the frequencies found for the K lines of highest frequency.

The critical absorption frequencies therefore give valuable information about the energy levels in the atom.

If W_0 denotes the energy of a normal neutral atom which has all its energy levels occupied by electrons, and W_m the energy of the atom when one m-quantum level is vacant, then

$$W_m = W_0 + E_m.$$

If now an electron drops from an n-quantum level to the vacant m-quantum level the energy of the atom diminishes by

$$W_n - W_m = (W_0 + E_n) - (W_0 + E_m) = E_n - E_m = h\nu,$$

where ν is the frequency of the radiation emitted. Since we are only concerned with energy differences, we may if we like take W_0 to be zero and the energy of an atom with one m-quantum level vacant as equal to E_m, the work required to move an electron from an m-quantum level in the normal atom to just outside the atom. It is found that there is only one one-quantum or K level, but that there may be three L levels, five M levels, seven N levels, five O levels, and three P levels. These levels are found from the critical absorption frequencies when these have been observed, or from the frequencies of the lines in the different series. It is supposed, for example, that the three L levels correspond to electron orbits of different shapes but about the same size. The three L levels are numbered 2(1, 1), 2(2, 1), 2(2, 2) respectively. The electrons in the normal atom are supposed therefore to be arranged in groups or sets corresponding to the groups of levels. This question is discussed at p. 107.

12. Energy of Electrons and Frequency of Rays. De Broglie's Apparatus.

As we have seen, an element may be caused to emit its characteristic X-ray spectrum by bombarding it with electrons, as when it forms the anti-cathode of a Coolidge tube. The bombarding electrons do not excite a particular spectral line unless they have enough energy. For example to cause the anti-cathode to emit its K series lines the bombarding electrons must have at least as much energy as is required to remove an electron from the K level outside the atom. When the K level is vacant, electrons may fall into it from any of the higher levels, so that all the K lines appear together as the energy of the bombarding electrons is gradually increased. If the potential difference between the cathode and anti-cathode is V, the energy of the bombarding electrons is Ve, so that to excite a series of lines due to electrons falling into a vacant n-quantum level requires that $Ve > E_n$. It is found that these conclusions from the theory are in agreement with the facts.

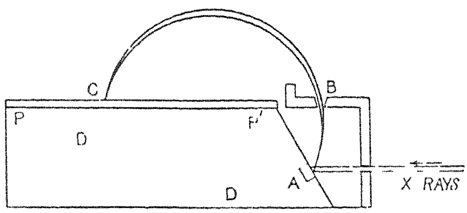


Fig. 17

The anti-cathode always emits a continuous X-ray spectrum as well as the spectral lines. It is found that this continuous spectrum ends sharply at a definite frequency, and no rays with a greater frequency than this limit are observed. The frequency at the limit is given by $Ve = h\nu$. This equation was shown to be accurately true by Duane and Hunt for values of V from 1500 to 170,000 volts.

When X-rays are passed through matter they cause the atoms to emit electrons. A quantum $h\nu$ of the incident rays is absorbed by an electron. If E_n is the energy required to remove the electron just outside the atom from its energy level, the electron is shot out of the atom with kinetic energy equal to $h\nu - E_n$. By measuring the kinetic energies of the electrons shot out by rays of known frequencies the energy levels of the atoms can be determined. The results obtained in this way agree with those obtained by the other methods.

The apparatus used by De Broglie for measuring the energies of the electrons emitted by matter when acted on by X-rays is shown in fig. 17. DD is a heavy metal block into which a small piece of the substance to be investigated is put at A. A narrow beam of X-rays is allowed to fall on the substance at A as shown. Some of the electrons from A pass through a slit at B and are deflected by a uniform magnetic field perpendicular to the plane of the paper so that they describe

circular paths and fall on a photographic plate PP′ at C. The whole apparatus is contained in a box which is exhausted, and the electrons are consequently not deflected by collisions with gas molecules. If m is the mass of an electron and v its velocity, then the radius R of its circular path is given by $\frac{mv^2}{R} = Hev$, where H is the strength of the magnetic field. We have also $m = \frac{m_0}{\sqrt{1 - v^2/c^2}}$, where m_0 is the mass when $v = 0$. By means of these equations v can be found with the help of the known value of e/m_0. The kinetic energy of the rays is equal to

$$c^2 m_0 \left\{ \frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right\},$$

and so can be calculated.

It is found that a series of lines is obtained on the photographic plate, showing that the electrons come off with definite velocities in accordance with the quantum theory. The energies of the electrons are always equal to $h\nu - E_n$, where ν is a frequency of the incident X-rays and E_n is the energy required to remove an electron from one of the energy levels in the atoms of the substance at A.

13. Scattering of X-rays. Classical Theory.

When a narrow beam of X-rays is passed through a thin plate of a substance of small atomic weight, in which the rays do not excite the characteristic X-rays of the substance, scattered X-rays are emitted by the plate in all directions. The intensity of these scattered rays in any direction can be measured by passing a narrow beam of them into an ionization chamber and measuring the conductivity they produce in it

It is found that the intensity of the scattered rays varies with the angle between their direction and that of the incident beam It is greatest when this angle is small, least at 90°, and greater near 180° than at 90°. If the intensity at 90° is taken as unity then that at 180° is about 2 and near 0° about 6.

If we assume the radiation to be scattered by free electrons according to the classical electromagnetic theory, it is easy to calculate the total energy of the scattered rays. On the classical theory (p 15), an electron moving with an acceleration f radiates energy at the rate of $\frac{1}{6\pi}\frac{e^2 f^2}{c^3}$ ergs per second. If F denotes the electric field strength in the incident X-rays then $mf = Fe$ where m is the mass and e the charge of the electron. The rate of radiation is therefore $\frac{e^4 F^2}{6\pi c^3 m^2}$. The energy density in the incident rays is $\frac{1}{2}(F^2 + H^2)$, where H is the strength of the magnetic field. But $F = H$, and the energy density is therefore F^2.

Hence the energy flowing through unit area in unit time is $c\overline{F^2}$, where $\overline{F^2}$ denotes the average value of F^2 in the incident rays, and the energy scattered by the electron in unit time is $\frac{e^4\overline{F^2}}{6\pi c^3 m^2}$.

If a piece of matter containing N electrons is placed in the incident beam, the total energy of the rays scattered by it will be equal to $\frac{Ne^4\overline{F^2}}{6\pi c^3 m^2}$. The ratio of the scattered energy to the energy flowing through unit area in the incident beam is therefore $\frac{Ne^4}{6\pi m^2 c^4}$. This ratio can be determined approximately by measuring the ionizations produced by the incident and scattered rays, and thus N can be found. In this way Barkla showed that the number of electrons in atoms of small atomic weight like carbon is not far from the atomic number. The classical theory thus appears to give results for light elements which agree at any rate roughly with the observed scattering.

14. A. H. Compton's Quantum Theory of Scattering.

It is found that the wave-length of the scattered rays is not equal to that of the incident rays, as it should be according to the classical theory. This change of wave-length by scattering has been investigated by A. H. Compton, who has shown that it can be explained on the quantum theory in a simple way.

We suppose that the electron receives a quantum of energy $h\nu$ from the incident rays and emits a quantum $h\nu'$ of scattered rays in a direction making an angle θ with the direction of the incident rays. The momentum of the incident quantum is $h\nu/c$ and that of the scattered quantum is $h\nu'/c$, so that the electron receives an impulse and starts off with a velocity v in a direction making an angle ϕ with the direction of the incident rays. Compton supposes that energy and momentum are conserved, so that

$$h\nu = h\nu' + mc^2\left\{\frac{1}{\sqrt{1-v^2/c^2}} - 1\right\},$$

$$\frac{h\nu}{c} = \frac{h\nu'}{c}\cos\theta + \cos\phi \cdot \frac{mv}{\sqrt{1-v^2/c^2}},$$

$$0 = \frac{h\nu'}{c}\sin\theta - \sin\phi \cdot \frac{mv}{\sqrt{1-v^2/c^2}},$$

where m is the mass of an electron at rest. Thus we have three equations to determine ν', v, and ϕ. Eliminating ϕ from the last two equations we get

$$\nu^2 - 2\nu\nu'\cos\theta + \nu'^2 = \frac{m^2c^2v^2}{h^2(1-v^2/c^2)},$$

which with the first equation gives

$$\nu' = \nu/\{1 + \alpha(1 - \cos\theta)\}$$

where $\alpha = h\nu/mc^2$.

If λ is the wave-length of the incident rays and λ' that of the scattered rays, so that $\nu\lambda = \nu'\lambda' = c$, then

$$\frac{\lambda'}{\lambda} = 1 + \alpha(1 - \cos\theta)$$

or

$$\lambda' - \lambda = (h/mc)(1 - \cos\theta).$$

Thus this theory indicates that λ' is greater than λ unless $\theta = 0$. The difference $\lambda' - \lambda$ is independent of λ, and the percentage change is therefore large for small wave-lengths but small for large ones. Putting in the known values of h, m, and c we get

$$\lambda' - \lambda = 0{\cdot}0242(1 - \cos\theta),$$

where $\lambda' - \lambda$ is in Ångstrom units ($1 = 10^{-8}$ cm.). Solving the equations for the angle ϕ, we get

$$\tan\phi = \frac{\cot(\theta/2)}{1 + \alpha},$$

and for the kinetic energy of the electron we get

$$2mc^2\alpha^2\cos^2\phi/(1 + 2\alpha + \alpha^2\sin^2\phi).$$

The electrons which receive energy from the X-rays in this way are called recoil electrons to distinguish them from photo-electrons, which receive energy $h\nu$.

A. H. Compton in 1923 scattered X-rays from a molybdenum anticathode by a carbon plate and measured the wave-lengths of the scattered rays with a Bragg spectrometer. It was found that the lines in the spectrum of the scattered rays were usually doublets, each doublet consisting of an unmodified line of wave-length λ and a modified line of wave-length λ'. The difference $\lambda' - \lambda$ was approximately equal to $0{\cdot}0242(1 - \cos\theta)$ in agreement with the quantum theory. This result has since been confirmed by several other observers.

15. Experiments supporting Quantum Theory of X-rays.

The quantum theory of the scattering of X-rays seems to require that the X-rays consist of particles having energy $h\nu$ and momentum $h\nu/c$. We may suppose that these particles execute some kind of oscillation of frequency ν as they move along with the velocity of light c. It is difficult to see how phenomena such as diffraction of X rays by crystals, which apparently depend on interference between trains of waves, can be explained on this theory.

Some very interesting experiments have been made by Bothe and Geiger, and by Compton and Simon, which strongly support the particle or quantum theory of X rays. Bothe and Geiger passed a narrow beam of X-rays through hydrogen. Two ionization chambers were arranged on either side of the beam in such a way that if a scattered quantum entered one chamber the scattering electron according to the theory would enter the other. The ionization in the chambers was greatly increased by ionization by collisions so that the effect due to a single quantum or electron could be detected. According to the theory we should expect that when ionization was produced in one chamber ionization would be produced at the same instant in the other. It was found that simultaneous ionizations occurred in about 10 per cent of the observed cases, which Bothe and Geiger considered to be considerably more than could be explained by chance coincidences.

Compton and Simon passed a narrow beam of X-rays of very short wave length into a C. T. R. Wilson cloud chamber, and examined the electron tracks produced. The recoil electrons produce short tracks and the photo electrons or β rays emitted by atoms with energy approximately $h\nu$ produce much longer tracks, so that it is easy to distinguish between the two. It was found that occasionally a recoil track started from the path of the X-ray beam and simultaneously a β-ray track from a neighbouring point outside the beam. The β-ray track was assumed to be produced by an electron shot out of an atom by the scattered quantum which produced the recoil track. The angle ϕ between the initial direction of the recoil track and the X-rays was observed, and also the angle θ between the X-ray beam and the direction from the beginning of the recoil track to the beginning of the β-ray track. It was found that the values of ϕ observed agreed approximately with those calculated by means of the equation

$$\tan\phi = \frac{\cot(\theta/2)}{1+\alpha}.$$

These results and those of Bothe and Geiger seem to support very strongly the particle theory of X-rays and to be quite inexplicable on the wave theory. It appears that there is a large number of facts which agree with the classical wave theory of light and X-rays and also a large number of facts which agree with a particle or quantum theory. Each set of facts agrees with one theory and seems definitely inconsistent with the other theory. Since agreement between a theory and a set of facts does not prove the theory true, whereas disagreement does prove the theory untrue or at least inadequate, the proper conclusion to be drawn is clearly that both theories are inadequate or untrue. Presumably a new theory will eventually be discovered which will be capable of explaining all the facts.

16. γ-rays.

The most penetrating rays emitted by radioactive bodies are called γ-rays and are found to have properties similar to those of very penetrating X-rays. They are not deflected by electric or magnetic fields and carry no charge of electricity. They affect a photographic plate and ionize gases like X-rays. The γ-rays from 30 mgm. of radium can be detected, after passing through 30 cm. of iron, by the ionization they produce. γ-rays are only emitted by radioactive bodies which also emit β-rays. This suggests that γ-rays are produced by β-rays, or vice versa. It has been shown that β-rays produce γ-rays when they are stopped by lead, just as X-rays are produced by the impact of high velocity electrons on the anti-cathode of an X-ray tube. It is also found that γ-rays produce β-rays when they are absorbed by matter just as X-rays cause the emission of β-rays. Rutherford showed that when β-rays are produced by γ-rays which were produced by β-rays from a radioactive body, then the β-rays produced by the γ-rays have the same energy as the β-rays which produced the γ-rays.

The spectra of the γ-rays from radioactive bodies have been examined by means of the Bragg crystal spectrometer or similar apparatus, and it is found that the spectra contain lines having wave-lengths characteristic of the elements emitting the rays. The wave-lengths range from $0{\cdot}07 \times 10^{-8}$ cm. to $0{\cdot}4 \times 10^{-8}$ cm. The γ-rays from some radioactive bodies appear to belong to the K or L series of X-rays, and so must be due to an electron falling from an outer level in the atom into the K or L level. The β-rays emitted by the radioactive nucleus may be supposed to knock an electron out of the K or L level, and the γ-rays to be emitted when an electron falls into the place left vacant.

When γ-rays are scattered by light elements like carbon, scattered rays having a longer wave-length are obtained as with X-rays. The change of wave-length appears to agree with that calculated on Compton's theory of X-ray scattering.

The γ-rays emitted by the nucleus may give their energy to the extra-nuclear electrons, as in the photo-electric effect, so that the atom emits electrons with definite energies. These secondary electrons, knocked out by the γ-rays from the nucleus, may be studied with the apparatus shown in fig. 17 on p. 168. A fine wire coated with the radioactive body is placed at A, and the electrons are focussed by the magnetic field on the photographic plate PP′.

It is found that some radioactive bodies give a many lined spectrum of secondary electrons on the plate. The energies of the electrons for a few of the most intense lines are given in the following table. Those electrons with discrete energies are often called secondary β-rays.

Element.	Energies in Electron Volts.
RaB	0·37, 1·50, 2·036, 2·600.
RaC′	5·14, 13·23, 16·67, 21·04.
RdAc	0·425, 0·454, 0·567, 1·305, 1·501.
AcX	0·455, 0·559, 1·708.

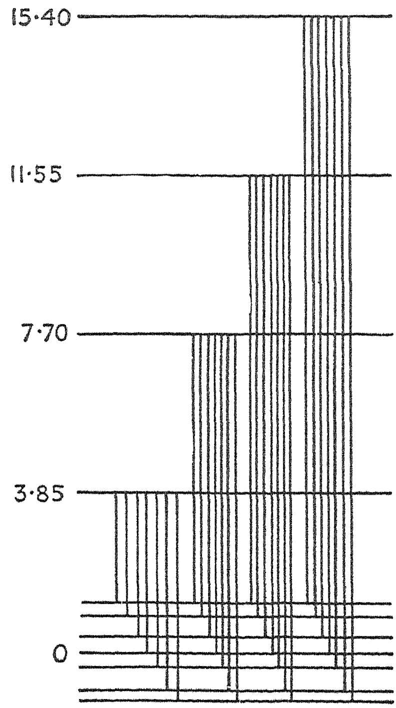


Fig. 18

If the γ-ray from the nucleus has energy E_γ and the energy it gives to the electronic system is E_S, then the energy of the electron E_β is given by $E_\beta = E_\gamma - E_S$. Frequently the γ-ray appears to merely knock a *K*, *L* or *M* electron out, so that for example $E_\beta = E_\gamma - E_K$, where E_K is the energy required to remove a *K* electron from the atom. The energies E_K, E_L, E_M, . . . are known accurately from the X-ray energies, so that the energy of the γ-rays may be deduced from the observed secondary electron energies.

It is found that there are a great many pairs of secondary β-ray and γ-ray energies with equal sums. For example, the secondary β-rays of radium C′ give the following pairs. The unit is 10^5 electron volts.

$$
\begin{aligned}
5{\cdot}14 + 21{\cdot}81 &= 26{\cdot}95\\
8{\cdot}98 + 17{\cdot}89 &= 26{\cdot}87\\
12{\cdot}86 + 14{\cdot}11 &= 26{\cdot}97\\
16{\cdot}67 + 10{\cdot}28 &= 26{\cdot}95
\end{aligned}
$$

$$
\begin{aligned}
5{\cdot}14 + 14{\cdot}11 &= 19{\cdot}25\\
8{\cdot}98 + 10{\cdot}28 &= 19{\cdot}26\\
16{\cdot}67 + 2{\cdot}54 &= 19{\cdot}21
\end{aligned}
$$

$$
\begin{aligned}
5{\cdot}14 + 10{\cdot}28 &= 15{\cdot}42\\
12{\cdot}86 + 2{\cdot}54 &= 15{\cdot}40
\end{aligned}
$$

The sums of these pairs are found to be approximately equal to multiples of 3·85. Thus $3{\cdot}85 \times 7 = 26{\cdot}95$, $3{\cdot}85 \times 5 = 19{\cdot}25$, and $3{\cdot}85 \times 4 = 15{\cdot}40$. It appears, therefore, that the γ-ray and secondary β-ray energies are equal to $3{\cdot}85n \pm C_m$, where n is a positive integer

and C_m a constant which has the same value for a number of the ray energies. Thus C_m is equal to 1·29 for all the energies in the list of pairs just given. These facts suggest the energy levels and transitions shown in fig. 18.

It was suggested by the writer that the energy levels which are multiples of 3·85 belong to the nucleus and that the small levels $\pm C_m$ represent energy received from or given to the electronic system. Thus a ray with energy $3{\cdot}85n - C_m$ may be due to a γ-ray with energy $3{\cdot}85n$ emitted by the nucleus which has given energy C_m to the electronic system or has given all its energy to an electron which has then

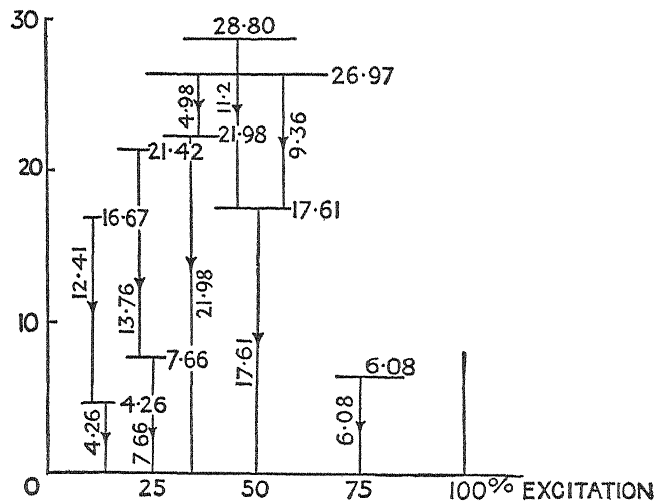


Fig 19

given energy C_m to the other electrons. It may be, however, that the levels $\pm C_m$ really belong to the nucleus.

Other systems of energy levels have been proposed. For example, the levels shown in fig. 19, due to Rutherford, Lewis and Bowden, give the principal γ-ray energies of RaC′.

The energies of the three most intense secondary beta rays from radium B have been very accurately measured by Scott and Rogers. The magnetic field was produced by a permanent cobalt steel magnet, and was measured to one part in 10,000 with an improved form of Cotton balance. The energies found in electron volts are $1{\cdot}513 \times 10^5$, $2{\cdot}045 \times 10^5$ and $2{\cdot}612 \times 10^5$.

17. The J Phenomenon.

The scattering of X-rays has been studied very extensively by Barkla, who finds results which apparently are not in agreement with Compton's quantum theory. The following summary of Barkla's results is quoted from one of his recent papers. (1) When a heterogeneous X-radiation is scattered, the scattered radiation has either precisely the same absorbability as the primary, or there is a well-marked difference between the absorbabilities when measured in any one substance. (2) When scattered radiation is different from the primary when measured in certain substances, its absorbability may be, and frequently is, precisely like that of the primary when measured in certain other substances. (3) Even after transmission through substances which show differences between primary and secondary radiations, there is still no difference between the two when measured in certain other substances. (4) When the scattered radiation is observed in different directions making angles of 30°, 60°, and 90° with the primary beam, and when these radiations differ from the primary in their absorbability in a certain substance, they in general differ by precisely the same amount. When a difference appears with increasing angle of scattering, it takes place abruptly by a jump.

The changes in absorbability are referred to by Barkla as the J phenomenon. He considers that a heterogeneous beam of X-rays has properties depending on the distribution of energy in its spectrum which cannot be deduced from the properties of the beams of homogeneous rays of which the heterogeneous beam may be regarded as composed. He considers that his results do not agree with either the classical wave theory or the quantum particle theory of X-rays.

Barkla's results seem to show that absorbability does not define the properties of a heterogeneous beam of X-rays. The absorbabilities of two beams may be the same in some different substances and different in other substances.

18. Cosmic Rays.

It has been known for a long time that air in a closed vessel is slightly ionized even in the absence of any radioactive or other source of ionizing radiations. Most of this ionization disappears when the closed vessel is surrounded by thick screens of lead or water, and so it was supposed to be due to traces of radioactive bodies in the ground or other bodies in the neighbourhood.

During the years 1911–14 Hess and Kolhoerster found that the ionization increases when the closed vessel is carried up in a balloon to high altitudes. At 9000 metres it became about forty times greater

than at sea level. Kolhoerster suggested that this increase must be due to penetrating radiation falling on the earth from outside which is absorbed as it passes down through the air.

In recent years this penetrating radiation has been carefully studied by many investigators, notably Millikan and A. H. Compton in America, and by Clay, Kolhoerster and Rossi in Europe. The radiation is now usually called *cosmic rays*. The ionization in closed vessels due to the cosmic rays, at sea-level, has been measured at many points all over the earth's surface, and it is found to depend on the earth's magnetic field. The magnetic equator where the earth's field is horizontal does not coincide with the geographical equator, but is approximately half-way between the magnetic poles. The magnetic latitude varies from zero at the magnetic equator to 90° at the magnetic poles and is determined by the earth's field.

It is found that the cosmic ray sea-level intensity is practically constant from magnetic latitude 50° to 90°, but drops from 1·7 ions per cm.^{3} per sec. at 50° to 1·5 ions per cm.^{3} per sec. at the magnetic equator. The value at 25° is 1·51. At high altitudes the latitude effect is much greater than at sea-level, and at the highest altitudes it has been stated that the intensity at the equator is less than 5 per cent of that at the poles. The lowest intensity at sea level occurs in Java, where the earth's magnetic field is greatest. There is a small longitude effect corresponding to the variation of the magnetic field with longitude.

These results show clearly that the cosmic rays which enter the earth's atmosphere from outside are charged particles, because uncharged particles such as photons or neutrons would not be affected by the earth's magnetic field.

In order that an electron or proton may be able to penetrate the earth's atmosphere and get to the earth's surface it must have about 3×10^9 electron volts energy. At latitudes greater than 50° electrons with this energy are not sufficiently deflected by the earth's magnetic field to be prevented from reaching the top of the atmosphere. At latitudes less than 50° some of the electrons with energy only $3 \times 10^9 EV$ are deflected away from the earth by its magnetic field, so the ionization at sea-level is less. At the equator only electrons with about $20 \times 10^9 EV$ can get to the earth. The drop from 1·7 ions per cm.^{3} at 50° to 1·5 at the equator is therefore due to the elimination of the primary rays, with energies between 3×10^9 and $20 \times 10^9 EV$, by the earth's field. It appears, therefore, that a large fraction of the primary rays have energies greater than $20 \times 10^9 EV$.

The variation of the ionization due to cosmic rays in a closed chamber with height above the earth's surface has been investigated by means of observations on mountains and in balloons and aeroplanes, and also with automatic recording instruments carried by

small unmanned balloons. Fig. 20 shows the relation between ionization and barometric pressure at about 45° magnetic latitude. The ionization increases rapidly with elevation until the pressure is about 15 cm. of mercury and then falls rapidly. It is supposed that the primary rays which come in from outside produce secondary rays by collisions with the air molecules, so that the maximum ionization due to the primary and secondary rays is not attained until the primary rays have passed through a considerable amount of air. The absorption of the primary and secondary rays by the air then causes the ionization to fall with decreasing altitude.

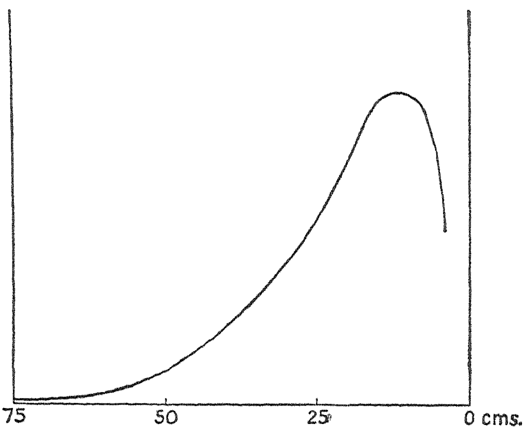


Fig. 20

The cosmic rays may be detected with a Geiger counter, which consists of a fine wire mounted along the axis of a thin metal cylinder. The cylinder and wire are usually put in a glass tube containing air or argon at a few centimetres pressure. The cylinder is charged negatively to about 1000 volts, and the wire is grounded through a very high resistance. The potential difference is adjusted so that when an ionizing particle passes across the cylinder a momentary discharge between the wire and cylinder takes place. The potential difference must be only slightly too small to produce a continuous discharge. The wire is connected through an amplifier to an oscillograph or an impulse counter so that the particles which pass through the counter are recorded or counted. In this way it is found that there is about one particle per minute per cm.^{2} at sea-level.

If two Geiger counters are placed with their axes parallel at a distance of, say, 50 cm. apart and are connected to an amplifier and impulse counter so that only simultaneous discharges in the two counters are counted, the number of rays which pass through both counters can be determined. It is found that this number is greatest

when the axes of the two counters are horizontal and in a vertical plane, and that it falls off rapidly when the plane containing the counters is inclined to the vertical, showing that most of the particles come down nearly vertically. If thick plates of lead or other materials are put between the two counters, the number of coincident discharges is diminished, and the mass absorption coefficients determined in this way are about the same as those found by measuring the ionization in a closed vessel at different depths in water or at different heights in the atmosphere. The ionization (I) in a closed chamber diminishes with the thickness of matter traversed by the rays approximately according to the equation $I = I_0 e^{-\mu x}$, where x is the thickness traversed and μ the absorption coefficient. The mass absorption coefficient is μ/ρ, where ρ is the density of the matter, so that if M is the mass per cm.^{2} traversed then $I = I_0 e^{-\mu M/\rho}$. It is found that μ/ρ is not constant but diminishes as M increases, so that μ/ρ is about 5×10^{-3} cm.^{2} gm.$^{-1}$ in the atmosphere, but only about $0{\cdot}2 \times 10^{-3}$ at a depth of 230 metres of water.

19. Cosmic Ray Showers.

It was found by Rossi that if several Geiger counters are placed below a sheet of lead then simultaneous discharges of the counters occur, showing that the lead emits several particles simultaneously in different directions. It is supposed that a cosmic ray particle collides with an atom in the lead and

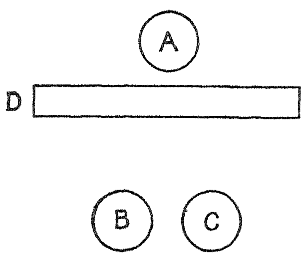


Fig. 21

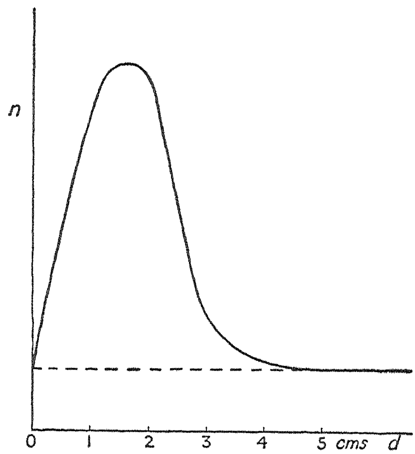


Fig. 22

causes it to emit several particles. The arrangement shown in fig. 21 was used by Sawyer to study this emission of showers of rays. Three Geiger counters ABC were connected through an amplifier to an impulse counter so that only simultaneous discharges of all three were counted. A plate of lead D was placed below the upper counter, and the number of coincident discharges was measured for different thicknesses of the lead plate. Fig. 22 shows the relation between the

number n of coincidences per hour and the lead thickness d. The dotted line shows the number of coincidences observed without the lead plate. These coincidences are accidental, due to rays from different bodies near.

The shower-producing rays (s.p.r.) are not the primary rays, but are secondary rays produced by the primary rays from the air or other material. The s.p.r. set off the counter A and enter the lead plate D in which they are absorbed, producing showers of rays which are also absorbed in the lead. If μ_1 is the absorption coefficient of the s.p.r. and μ_2 that of the shower rays in lead, then the number of showers (n) should vary with the thickness (d) of the lead plate according to the equation

$$n = A\{e^{-\mu_1 d} - e^{-\mu_2 d}\},$$

where A is a constant. The observed values of n with the accidental coincidences subtracted agree well with this expression, and the values of μ_1 and μ_2 can be obtained from them. The maximum occurs when $d = \frac{1}{\mu_2 - \mu_1} \log \frac{\mu_2}{\mu_1}$. This value of d was about 1·3 cm., and the values of μ_1 and μ_2 were 0·50 and 2·58.

If a block of lead, thick enough to absorb the air s.p.r. completely, is put above A, then the primary rays produce lead s.p.r. in the block, and these produce showers in the lead plate below A. It is found that the absorption coefficients of the s.p.r. produced in lead and other materials are different from the air s.p.r.

If plates of different materials are placed above the large lead block, the intensity of the primary rays is reduced and so the number of showers is reduced. In this way the absorption of the primary rays can be obtained. Sawyer found that the coefficient of absorption of the primary rays which produce the s.p.r. is considerably greater than the cosmic ray absorption coefficients as found from the ionization in closed vessels. This shows that the s.p.r. are produced by the more easily absorbed constituents of the cosmic rays. It is found that the number of showers increases with altitude much more rapidly than the ionization in closed vessels.

Hoffmann has observed extraordinarily large showers in closed vessels. These bursts of ionization occur much less frequently than ordinary small showers. The total energy to produce such bursts is sometimes as much as 10^{11} EV.

Very interesting results have been obtained by observing the tracks of the cosmic ray primary and secondary particles in C. T. R. Wilson cloud expansion chambers. Skobelzyn first observed straight nearly vertical tracks due to the cosmic rays. Mott-Smith and Locher placed a Geiger counter above the expansion chamber and showed that when a track was obtained, which when produced

passed through the counter, then the counter was discharged.

Anderson placed the cloud chamber in a strong magnetic field and measured the curvature of the primary cosmic ray particle tracks and also that of the shower particle tracks. He found that about half the tracks are bent as for negatively charged particles, and the other half as for positively charged particles. Moreover, the tracks of positively charged particles are thin tracks just like negative electron tracks, and not like the thicker tracks of protons and alpha rays. He therefore concluded that the cosmic ray tracks due to positively charged particles are due to particles with about the same mass as electrons. These new particles are called positrons Anderson found that the shower particles are about half positrons and half electrons Positrons have since been obtained from radioactive bodies, and it has been shown that e/m for positrons is equal to $-e/m$ for electrons. The discovery of positrons is the most interesting result so far obtained by the study of cosmic rays.

20. Mesons.

In 1936 Anderson and Neddermyer found a track made by a cosmic ray particle in a cloud chamber, in a magnetic field, which could not be explained as being due to either a proton or an electron. The product of the magnetic field strength H and the track radius of curvature ρ is proportional to the momentum of the particle. It was found that the new track had fewer droplets per centimetre than a proton track of the same curvature. This suggested that the particle had a higher velocity than a proton with the same momentum, and therefore a smaller mass. Many such tracks have since been observed and it is found that they are due to particles with masses about 200 times that of an electron. These particles are called mesons. Some mesons have a charge $+e$ and about as many a charge $-e$. At high altitudes the number of mesons is much greater than at sea level.

It is found that mesons are unstable and disintegrate into electrons and neutrinos or photons. The average half-life of a meson is about 2×10^{-6} second.

The primary cosmic rays which enter the earth's atmosphere from outside appear to be mainly protons with energies up to several thousand million electron volts. These protons collide with the nitrogen and oxygen atoms producing neutrons, mesons, electrons and photons.' All these particles produce other particles as they move down through the atmosphere. For example, a photon may produce a pair of electrons, one positive and one negative. Electrons knock electrons out of atoms and produce photons. The photons produce electrons and so on. At sea level cosmic rays consist of a very penetrating component, probably mainly protons, and much less penetrating components probably mainly electrons and photons.

References

1. *X-rays and Crystal Structure.* W. H. and W. L. Bragg.
2. *X-rays and Electrons.* A. H. Compton.

CHAPTER VIII

Optical Spectra

1. Principal, Sharp, Diffuse, and Fundamental Series.

The spectra of atomic hydrogen and of ionized helium are discussed in the chapter on quantum theory, and X-ray spectra in the chapter on X-rays. In the present chapter the series of spectral lines found in the optical spectra of many elements will be considered.

About 1890 Kayser and Runge discovered series of lines in the spectra of the alkali metals, and soon after this Rydberg showed that the frequencies [x] of these lines could be expressed approximately as the differences of two *terms*, each term being of the form

$$\frac{N}{(n+\alpha)^2},$$

where N is a constant having the same value in all the series, α a small constant, and n an integer. Rydberg's formula is analogous to Balmer's formula for the frequencies in the spectrum of atomic hydrogen, which are equal to the differences between two terms each of which is N/n^2.

Ritz showed that a more exact expression for the frequencies is got by taking the difference between two terms of the form $\dfrac{N}{(n+\alpha+\beta/n)^2}$ or $\dfrac{N}{(n+\alpha+\beta/n^2)^2}$, where α and β are small constants.

A particular *series* is represented by

$$\bar{\nu} = N\left\{\frac{1}{(n'+\alpha)^2} - \frac{1}{(n+\alpha)^2}\right\},$$

where n' has the same integral value in all the lines of the series and $n = n'+1$, $n'+2$, $n'+3$, &c. When n is very large $(n+\alpha)^{-2} = 0$, and we get as the limit of the series $N/(n'+\alpha)^2$. The lines in a series

[x] In practical spectroscopy, the word *frequency* is habitually used in the sense of "number of wave-lengths per centimetre". The word *wave-number* is sometimes used with the same meaning. To prevent the possibility of confusion, frequency in this conventional sense will be denoted here by the symbol $\bar{\nu}$, and frequency in the strict sense by ν as hitherto. The constant N, called the Rydberg constant, has the value 109732 cm.$^{-1}$. The constant $2\pi^2me^4/h^3$ of Chap. V, §14, is nearly equal to Nc and $\nu = c\bar{\nu}$.

get fainter as the limit is approached. Four such series have been found in the arc spectra of the alkali metals and alkaline earth metals. These series are known as the principal series, the sharp series, the diffuse series, and the fundamental series. The formulæ usually used for the frequencies in these four series are.

Principal·	$P(n) = P(\infty) - N/(n+P)^2$,	$n = 2, 3, 4 \ldots$,
Sharp:	$S(n) = S(\infty) - N/(n+S)^2$,	$n = 2, 3, 4 \ldots$,
Diffuse:	$D(n) = D(\infty) - N/(n+D)^2$,	$n = 3, 4, 5 \ldots$,
Fundamental.	$F(n) = F(\infty) - N/(n+F)^2$,	$n = 4, 5, 6 \ldots$,

where e.g. $P(n)$ is the frequency of a line in the P series, and P, S, D, F are the values of the constant a in the respective series. In many cases each line of a series is not a single line but a doublet or triplet. Each series according to the above formulæ is determined by two constants, such as $P(\infty)$ and P, which are different in different series. It was shown by Rydberg that these constants for different series are related in a simple way. He found that

$$S(\infty) = D(\infty) = \frac{N}{(2+P)^2},$$

and also that

$$P(\infty) = \frac{N}{(1+S)^2}.$$

From this it follows that $P(\infty) - S(\infty)$ is equal to

$$P(\infty) - \frac{N}{(2+P)^2} = P(2).$$

Runge pointed out that $D(\infty) - F(\infty) = D(3)$,

or

$$F(\infty) = \frac{N}{(3+D)^2}.$$

The four series may therefore be represented by

$$P(n) = \frac{N}{(1+S)^2} - \frac{N}{(n+P)^2}, \quad n = 2, 3, 4 \ldots,$$

$$S(n) = \frac{N}{(2+P)^2} - \frac{N}{(n+S)^2}, \quad n = 2, 3, 4 \ldots,$$

$$D(n) = \frac{N}{(2+P)^2} - \frac{N}{(n+D)^2}, \quad n = 3, 4, 5 \ldots,$$

$$F(n) = \frac{N}{(3+D)^2} - \frac{N}{(n+F)^2}, \quad n = 4, 5, 6 \ldots,$$

so that only four constants S, P, D, and F are required besides N.

These formulæ are usually written in an abbreviated form as follows:

$$\begin{aligned} P(n) &= 1S - nP, \\ S(n) &= 2P - nS, \\ D(n) &= 2P - nD, \\ F(n) &= 3D - nF. \end{aligned}$$

Thus $D(3)$ means the spectral line in the D series for which $n = 3$, but $3D$ stands for the term $\dfrac{N}{(3+D)^2}$.

2. The Combination Principle.

Rydberg and Ritz pointed out that spectral lines with frequencies given by the differences between any possible pair of the terms nP, nS, nD, and nF might be expected to appear. For example, $2S - nP$, $3S - nP$, and so on. A great many such lines have been discovered, so that according to this principle, known as the combination principle, lines may occur with frequencies given by the abbreviated expression $nX - n'X'$, where n and n' are integers, and X and X' stand for P, S, D, or F. This result may be compared with the formula for the frequencies of the lines in the spectrum of atomic hydrogen, $\bar{\nu} = \dfrac{N}{n^2} - \dfrac{N}{n'^2}$, where n and n' are integers.

It appears as though spectral lines could be arranged in an infinite number of series each having an infinite number of terms. Of course only a comparatively small number of the lines indicated by these formulæ have been actually observed.

According to the quantum theory the terms like nP, nS, &c., represent energy levels in the atoms. When an electron falls from one energy level to another having less energy, radiation is emitted of frequency ν given by $h\nu = E_1 - E_2$. The line $P(3)$ given by $P(3) = \dfrac{N}{(1+S)^2} - \dfrac{N}{(3+P)^2}$, for example, is supposed to be emitted when an atom changes from a stationary state having energy $C - \dfrac{Nhc}{(3+P)^2}$, where C is a constant, to another state having energy $C - \dfrac{Nhc}{(1+S)^2}$, so that the frequency of the radiation emitted is given by

$$\bar{\nu} = P(3) = \frac{E_1 - E_2}{hc} = \frac{C}{hc} - \frac{N}{(3+P)^2} - \left(\frac{C}{hc} - \frac{N}{(1+S)^2}\right),$$

or

$$\bar{\nu} = P(3) = \frac{N}{(1+S)^2} - \frac{N}{(3+P)^2}.$$

3. Spectra of Atoms which have lost Electrons.

By using sparks between electrodes very near together in a very perfect vacuum and working with very large potential differences, Millikan and Bowen have obtained the spectra of many atoms which have lost one or more electrons, thus extending work of this kind initiated by Fowler, Paschen, and others.

We should expect atoms containing the same number of electrons but different nuclear charges to give similar spectra. Increasing the nuclear charge must diminish the distance of the electrons from the nucleus and so increase the energy of the atom, but it should not alter the arrangement of the electrons or the relative sizes of their orbits. Thus increasing the nuclear charge, keeping the number of electrons in the atom constant, should shift the whole spectrum towards shorter wave-lengths without altering the arrangement of the lines in it. In the case of atoms having only one electron we should expect the frequencies to be given by

$$\bar{\nu} = Z^2 N\left(\frac{1}{n^2} - \frac{1}{n'^2}\right),$$

where Z is equal to the number of ionic charges in the nucleus, or the atomic number, as in the cases of atomic hydrogen and ionized helium.

4. Spectra of Atoms containing 1, 2, 3, and 4 Electrons.

If we indicate the number of electrons lost by an atom by a suffix, so that, for example, O_{III} indicates an oxygen atom which has lost three electrons, then the atoms

$$H,\ He_I,\ Li_{II},\ Be_{III},\ B_{IV},\ C_V,\ N_{VI},\ O_{VII},\ F_{VIII},$$

which all contain only *one* electron, should all give spectra represented by the above equation with $Z = 1$ for H, $Z = 2$ for He_I, $Z = 3$ for Li_{II}, and so on. Lines given by $\bar{\nu} = Z^2 N\left(\frac{1}{n^2} - \frac{1}{n'^2}\right)$ have so far only been observed in the spectra of H, He, and Li.

The following atoms all contain *two* electrons: He, Li_I, Be_{II}, B_{III}, C_{IV}, N_V, O_{VI}, F_{VII}. The spectra of these atoms should contain series represented approximately by

$$\bar{\nu} = Z^2 N\left\{\frac{1}{(n' + a')^2} - \frac{1}{(n + a)^2}\right\},$$

where Z is equal to the atomic number, since such series have been found in the spectra of helium and Be_{II}.

The following atoms all contain three electrons: Li, Be_I, B_{II}, C_{III}, N_{IV}, O_V, F_{VI}.

The spectrum of Li contains series represented by

$$\tilde{\nu} = \frac{N}{(n'+a')^2} - \frac{N}{(n+a)^2},$$

so we should expect similar series, with Z^2N instead of N, in the spectra of the other atoms with three electrons. Such series have been found by Fowler, Paschen, Millikan and Bowen, and others.

The following table given by Millikan and Bowen gives the values of the series terms for Li, Be_I, B_{II}, and C_{III}. The difference between two such terms for the same element gives the frequency of a spectral line.

	$n =$	2	3	4	5
	$109732/n^2 =$	27,433	12,192·8	6,858·1	4,389·1
P_1 terms	Li	28,582·5	12,560·4	7,018·2	4,473·6
	$Be_I \div 4$	28,736·3	12,596·2	7,030·1	4,477·6
	$B_{II} \div 9$	28,616·1			
	$C_{III} \div 16$	28,465·3	12,504·3		
D terms.	Li	—	12,203·1	6,863·5	4,389·6
	$Be_I \div 4$	—	12,206·9	6,865·1	4,393·7
	$B_{II} \div 9$	—	12,207·8		—
	$C_{III} \div 16$	—	12,208·3	—	—
F terms.	Li	—		6,856·1	4,381·8
	$Be_I \div 4$	—	—	6,858·8	4,389·5
	$B_{II} \div 9$	—	—	6,860·2	—

The second row of this table gives the values of $N = 109,732$ divided by $n^2 = 4$, 9, 16, or 25. These are the values of the terms in the equation $\tilde{\nu} = \frac{N}{n'^2} - \frac{N}{n^2}$, which gives the frequencies of the spectral lines of atomic hydrogen.

The remaining rows give the values of the series terms deduced from the observed frequencies of the lines of Li, Be, B, and C atoms. The terms attributed to Be_I are divided by 4, those attributed to B_{II} by 9, and those attributed to C_{III} by 16.

We see that the resulting quotients are nearly equal to the values of N/n^2 and are nearly the same for the different atoms. Thus, for example, the spectrum of B_{II} contains lines having nearly nine times the frequencies of lines in the spectrum of Li.

It is supposed that in these atoms, which contain three electrons, two electrons describe small orbits near the nucleus and the third a much larger orbit. The outer electron has a number of possible orbits

according to the Bohr theory, and lines with frequencies nearly equal to $Z^2N\left(\frac{1}{n'^2}-\frac{1}{n^2}\right)$ are emitted when the outer electron drops from one outer orbit to another farther in. The nucleus and the two inner electrons form a small group having a total charge $(Z-2)e$ where Z is the atomic number, so that so long as the orbits of the outer electron do not come near this group the orbits are very similar to those of a single electron moving near a nucleus with charge $(Z-2)e$.

For	Li, $Z=3$,	so that	$(Z-2)e=e$;
for	Be, $Z=4$,	so that	$(Z-2)e=2e$;
for	B, $Z=5$,	so that	$(Z-2)e=3e$,
and for	C, $Z=6$,	so that	$(Z-2)e=4e$.

Thus the appropriate constants for Li, Be_I, B_{II}, and C_{III} are N, $4N$, $9N$, and $16N$.

The following atoms all contain *four* electrons:

$$\text{Be},\ B_I,\ C_{II},\ N_{III},\ O_{IV},\ F_V.$$

These atoms have two outer electrons and should therefore give series represented by

$$\frac{Z^2N}{(n'+\alpha')^2}-\frac{Z^2N}{(n+\alpha)^2},$$

since such series are found in the spectra of Ca, Sr, Ba, and Ra, which all have two outer electrons.

5. Spectra of Atoms with 11 Electrons.

The following atoms all contain 11 electrons:

$$\text{Na},\ Mg_I,\ Al_{II},\ Si_{III},\ P_{IV},\ S_V.$$

Sodium is supposed to have 2 electrons very near the nucleus, 8 farther out, and 1 outer electron which describes orbits usually well outside the inner group consisting of the nucleus and 10 electrons. The electrons in the other atoms with 11 electrons should be arranged in the same way, so that we should expect all these atoms to give series containing lines represented by

$$Z^2N\left(\frac{1}{(n'+\alpha')^2}-\frac{1}{(n+\alpha)^2}\right).$$

This is found to be the case. The following table gives the term values for these atoms divided in each case by $(Z-10)^2$, Z being the atomic number.

	n =	3	4	5	6
	N/n^2 =	12,192·78	6,858·44	4,389·49	3,048·19
S	Na/1	41,449·0	15,709·5	8,248·3	5,077·3
	Mg/4	30,316·9	12,865·6	7,120·3	4,517·3
	Al/9	25,494·89	11,476·82	6,535·29	-
	Si/16	22,756·83	10,633·65	6,168·72	
	P/25	20,979·65	10,061·63	5,914·35	—
	S/36	19,729·56	9646·22		—
P	Na/1	24,475·7	11,176·1	6,406·3	4,151·3
	Mg/4	21,376·6	10,154·0	5,949·6	3,909·2
	Al/9	19,504·01	9,526·85	5,664·93	-
	Si/16	18,272·62	9,105·53	5,471·16	
	P/25	17,401·88	8,802·20	-	-
	S/36	16,753·64		—	-
D	Na/1	12,276·2	6,900·4	4,412·5	3,061·9
	Mg/4	12,441·3	6,988·8	4,461·6	3,091·6
	Al/9	12,611·0	7,074·3	4,508·72	3,119·96
	Si/16	12,733·65	7,131·84	4,539·22	
	P/25	12,811·8	7,164·01		-
	S/36	12,856·59	—		
	Na/1		6,860·4	4,390·4	3,043·0
	Mg/4		6,866·8	4,394·3	3,051·2
	Al/9	—	6,871·28	4,397·61	3,053·83
	Si/16	-	6,874·19	4,399·97	3,055·97
	P/25	—	6,876·38	4,401·46	—
	S/36	—	6,878·11	-	—
F''	Na/1	—	-	4,388·8	-
	Mg/4		-		3,048·7
	Al/9	—	-	4,391·80	3,050·30
	Si/16	-	—	4,392·31	3,050·84
	P/25	—	—	4,392·74	3,051·14
	S/36	—	—	4,393·32	-
F'''	Na/1	-	-	-	3,046·3
	Mg/4	-			
	Al/9	-	-	-	3,049·64
	Si/16			—	3,049·81
	P/25			—	3,050·21

This table is taken from a paper by Bowen and Millikan (*Physical Review*, March, 1925). The values of the series terms in it are due to Fowler Paschen, and Millikan and Bowen.

It will be seen that the quotients of the D and F terms are nearly equal to the corresponding values of N/n^2. This is interpreted, as before, to mean that the orbits of the outer electron involved are well outside the inner orbits. Similar results to the above have been obtained with several other sets of atoms. We may remark, for example,

that potassium can be made to give a spectrum very similar to that of argon. The normal K atom has 19 electrons and the argon atom 18. It is therefore supposed that the K spectrum which resembles that of argon is due to K atoms which have lost one electron.

6. Analogy with Moseley's Law. Second Quantum Numbers.

These results on optical spectra are closely analogous to Moseley's law for X-rays. Moseley found that the frequencies of the K_a lines in X-ray spectra are given by

$$\bar{\nu} = \left(1 - \frac{1}{2^2}\right) N \left(Z - a\right)^2,$$

so that $\sqrt{\bar{\nu}}$ is proportional to $Z - a$, where Z is the atomic number and a is a constant. The square roots of the values of the series terms of which the quotients by $(Z - 10)^2$ are given in the above table are nearly proportional to $Z - 10$, except in the cases of the terms near the top of the table. This result is analogous to Moseley's law with $a = 10$.

Millikan and Bowen have pointed out that Moseley's law applies to the optical spectra of many other similar series of atoms.

In the quantum mechanics theory of atoms containing only one electron, like hydrogen and ionized helium, it was found that the possible energies are equal to $-2\pi^2 mZ^2e^4/n^2h^2$ and that the atoms are degenerate, that is to say, there may be several different proper functions or solutions of Schrödinger's equation for each energy proper value, l and m having various values for a given n (p. 106).

The average value of the distance r between the nucleus and the electron for an atom with one electron and proper function w_n is given by

$$\bar{\bar{r}} = \int \bar{w}_n r w_n \, dx\, dy\, dz.$$

On evaluating this integral it is found that

$$\bar{\bar{r}} = \frac{an^2}{Z}\left\{1 + \frac{1}{2}\left\{1 - \frac{l(l+1)}{n^2}\right\}\right\},$$

where $a = h^2/4\pi^2 me^2$. Thus $\bar{\bar{r}}$ depends on the quantum numbers n and l. It increases rapidly with n and decreases as l increases, especially when n is small.

The following are some values of n, l and $\bar{r}$.

n	l	$\bar{\bar{r}}$	n	l	r
1	0	$\frac{3}{2}\frac{a}{Z}$	3	0	$\frac{27}{2}\frac{a}{Z}$
2	0	$6\frac{a}{Z}$	3	1	$\frac{25}{2}\frac{a}{Z}$
2	1	$5\frac{a}{Z}$	3	2	$\frac{21}{2}\frac{a}{Z}$

The atoms containing, for example, eleven electrons are supposed to have ten electrons near the nucleus, and one usually farther out which therefore moves in a field which at relatively great distances from the nucleus is roughly the same as that due to a charge $(Z - 10)e$. The possible energies of this electron are therefore roughly equal to the energies of an atom with a nucleus having a charge $(Z - 10)e$ and one electron. This is true when the electron is unlikely to be found near the nucleus and the group of ten electrons near it. When n is small $\bar{r}$ is small, so the possible energies differ considerably from $-2\pi^2 me^4(Z-10)^2/n^2h^2$. Also, since $\bar{r}$ depends on l as well as n, the possible energies also depend on l, so that the atom is not degenerate with respect to l but has different energies for each value of l with the same value of n. For example, with $n = 3$ we get three possible energies corresponding to the three values of l, 0, 1 and 2 instead of the single value $-2\pi^2 me^4(Z - 10)^2/9h^2$ for an atom with one electron and a nucleus with charge $(Z - 10)e$. In general, corresponding to any value of n there are n values of l and so n different possible energies.

The selection rule * for l is that it can only change by ± 1, so that, for example, if the electron is in a state with $n = 3$ and $l = 2$, then it cannot change to a state with $l = 0$. The quantum numbers used by different writers to designate the different possible energies, or series term values, are not always n and l. They may be $n + 1$ or $n + 2$ and $n_1 = l + 1$ for example.

If the potential V of the outer electron is equal to $(Z-10)e^2/r + f(r)$, where $f(r)$ is the effect of the ten inner electrons, then $f(r)$ will be small unless r is very small. We may regard $f(r)$ as a perturbation, so that the differences between the energies E_n and the values of $-2\pi^2 me^4(Z - 10)^2/n^2h^2$ will be determined by the matrix elements of $f(r)$ or $\int \bar{w}_n f(r) w_m dv$. The proper functions w_n depend on the quantum numbers l and n, and as l increases the value of w_n near the origin diminishes. Thus, since $f(r)$ is small, except near the origin, the energy changes due to the perturbation will diminish as l increases. The energies with $l = 0$ differ considerably from the hydrogen-like energies, but those with $l = 2$, or more, are nearly equal to hydrogen-like energies.

The series terms $S(n)$, $P(n)$, $D(n)$ and $F(n)$ are supposed to have $l = 0, 1, 2, 3$ respectively. The selection rule, that l must change by ± 1, then explains why series due to transitions between P and S, P and D, D and F terms, but not S and D or P and F terms, are observed.

The D and F terms are nearly equal to hydrogen-like terms, but the P and especially the S terms differ greatly from hydrogen-like terms.

* See p. 133.

7. The Series Terms for Sodium.

The values of the series terms for sodium are shown graphically in the accompanying figure due to Bohr (fig. 1). The distances of the black dots from the vertical line on the right are proportional to the values of the series terms. The S terms are in the top row, the P terms in the second row, and so on. The curved lines are drawn through terms having the same principal quantum number, and the distances of these lines from the vertical line on the line for $n_1 = 5$ are proportional to the corresponding terms of a hydrogen atom.

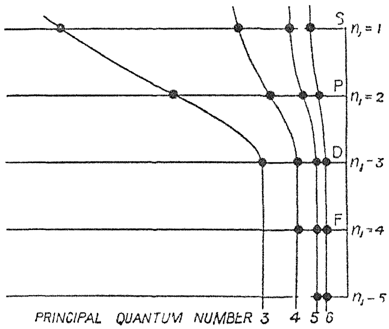


Fig 1

We see that the D, F, and $n_1 = 5$ terms are nearly equal to the hydrogen terms but the S and P terms are much larger. It is supposed that the S and P electron orbits penetrate into the group of 10 electrons around the nucleus and so are acted on by the strong field of the nucleus, which increases the energy of the orbits. All these terms are for a neutral sodium atom, and correspond to the different possible orbits of the outer electron. The first term of the principal series according to this diagram is due to a transition from the orbit with principal number 3 and second number 1 to the orbit with principal number 3 and the second number 2. This series is represented by $P(n) = 1S - nP$, so that in this case the principal quantum number is equal to $n + 1$ since $n = 2$ for the first line of the principal series.

The possible transitions, with emission of light, on the diagram are from any dot to another dot which lies to the left on the next line above or the next line below, and so corresponds to less energy, because the energy is proportional to a constant minus the term values.

If n is the principal and n_1 the second quantum number of an orbit or atomic energy level, then the level may be designated by n with a suffix n_1. Thus 3_1 denotes the lowest energy level of the neutral sodium atom according to the diagram.

8. Series Lines and Absorption.

The normal state of an atom is the state of smallest energy. An atom in a given state can only absorb light the quanta of which are equal to the quanta which the atom emits when it changes from another state of greater energy to its actual state. Thus, according to the quantum theory, sodium vapour at comparatively low temperatures at which all the atoms are in the normal 3_1 state should absorb light

with the frequencies of the lines of the principal series. This is found to be the case. When white light is passed through sodium vapour the absorption spectrum obtained shows the principal series lines only. The same thing is true of other elements. If the frequency of the incident light is greater than that of the limit of the principal series, the outer electron may be removed right out of the atom by a quantum of the incident light. The light is absorbed in this case and the vapour is ionized It is found that the absorption spectrum shows continuous absorption of frequencies greater than the limit of the principal series, exactly in agreement with this deduction from the quantum theory.

9. Doublets and Triplets.

The lines of many series are not single lines but doublets, triplets, or more complicated groups of lines The P terms are double or triple and the different values may be indicated by a suffix thus, mP_i, where $i = 1$ or 2 for a series of doublets, and $i = 1, 2,$ or 3 for a series of triplets. If the P terms are double, then the S and D series lines are doublets with constant frequency differences, since $S(n) = 2P - nS$ and $D(n) = 2P - nD$. The constant frequency difference is $2P_1 - 2P_2$ In this case the P series lines are doublets with frequency difference $nP_1 - nP_2$. This difference diminishes to zero as n increases, since both nP_1 and nP_2 tend to zero.

On the quantum mechanics theory the doubling of a term is supposed to be due to the series electron having two possible states with slightly different energies for each set of values of the principal quantum number n, the second quantum number l or n_1, and the quantum number m. It is supposed that the electron has angular momentum and a magnetic moment as if it was spinning about an axis through its centre. The component of the spin angular momentum along any axis like Oz may be either positive or negative with magnitude $h/4\pi$, and the magnetic moment component is $\pm eh/4\pi mc$. The two different states have quantum numbers n, l, m and a spin quantum number $s = \pm 1$.

According to Pauli's exclusion principle, not more than one electron in an atom can be in the same state or have the same set of quantum numbers n, l, m and s.

The lines of the principal series of the alkali metals are doublets. The well-known D lines of sodium are the first doublet in the sodium P series. Millikan and Bowen have observed the first doublet of the principal series of the atoms Li, Be_I, B_{II}, C_{III}, N_{IV}, and O_V, which all contain three electrons. If $\Delta\nu$ denotes the frequency difference of one of these doublets then they find that $\sqrt[4]{\Delta\nu/0{\cdot}365}$ is nearly equal to $Z - 2$ in each case. For example, for lithium $Z - 2 = 1$ and $\sqrt[4]{\Delta\nu/0{\cdot}365} = 0{\cdot}981$, and for oxygen $Z - 2 = 6$ and $\sqrt[4]{\Delta\nu/0{\cdot}365} = 6{\cdot}184$. According to Sommerfeld's theory of the fine structure of the spectral lines of H, He_I, Li_{II} discussed in the chapter on the quantum theory, $\Delta\nu$ for the H_α line, the first line in the Balmer series, is $0{\cdot}365$, and for other similar doublets should be proportional to the fourth power of the nuclear charge Ze. In the case of the atoms Li, Be_I, B_{II}, C_{III}, N_{IV}, and O_V the charge on the nucleus and group of two electrons near the nucleus is $(Z - 2)e$, so that on Sommerfeld's theory we should expect $\Delta\nu$ to be nearly equal to $0{\cdot}365(Z - 2)^4$, which is just what Millikan and Bowen find. Similar doublets with separations given by about $0{\cdot}365(Z - s)^4$, where s is about $3{\cdot}5$, have been found in the X-ray spectra of practically all the elements.

10. Band Spectra. Quantum Theory.

The spectra of many compounds and of diatomic elements contain groups of lines which are called bands. In these bands the lines get closer together from one side of the band to the other, so that one side of the band has a well-defined edge called the head of the band where the lines are very close together. Such spectra frequently contain a series of such bands, all with heads on the same side. A good example of such a band spectrum is the spectrum of the light from the inner cone of a Bunsen flame, which is usually attributed to carbon monoxide. Cyanogen and hydrochloric acid also give band spectra. Many elements give so-called many-lined spectra, which are probably band spectra in which the different bands overlap and so become confused. An example of such a spectrum is the many-lined spectrum of hydrogen, which is attributed to H_2. Several bands have been located in this spectrum by O. W. Richardson and others.

According to the quantum theory the bands are emitted by rotating molecules. The rotating molecules are supposed to be only able to exist in a number of definite states characterized by definite rotational energies, and to emit light when they change from one such state to another with less energy. The frequency of the light emitted is given by the usual equation $h\nu = E_1 - E_2$. Let K denote the moment of inertia of a molecule about its axis of rotation, and θ the angle through which it has turned. Then according to the quantum theory we have $\int_0^{2\pi} K\dot{\theta}d\theta = nh$, where n is an integer. Hence, since $\dot{\theta}$ is constant, $K\dot{\theta} = nh/2\pi$.

The kinetic energy of the molecule is therefore given by

$$E_n = \tfrac{1}{2}K\dot{\theta}^2 = \frac{1}{8\pi^2}\frac{h^2n^2}{K}.$$

We assume that n can only change by one unit at a time, so that when it is $+$ it can change to $n-1$, and when $-$ to $n+1$. The frequency of the light emitted is then

$$\nu = \frac{E_n - E_{n-1}}{h} = \frac{h}{8\pi^2K}\{n^2 - (n-1)^2\} = \frac{h}{8\pi^2K}(2n-1),$$

when n is $+$, and $\nu = -\frac{h}{8\pi^2K}(2n+1)$, when n is $-$.

The frequencies emitted are therefore

n		ν.
$+4$	....	$7\nu_0$
$+3$	....	$5\nu_0$
$+2$	. .	$3\nu_0$
$+1$	.	ν_0
0	....	$-\nu_0$
-1	....	ν_0
-2	....	$3\nu_0$
-4	...	$7\nu_0$

where $\nu_0 = h/8\pi^2K$. The frequency $-\nu_0$ for $n=0$ indicates an absorption of energy by the molecule. The spectrum should therefore consist of lines with frequencies ν_0, $3\nu_0$, $5\nu_0$, &c. Such spectra have been observed far in the infra-red, notably for water vapour, by Rubens and Bahr.

If we suppose that the molecule also vibrates with a frequency ν_1, then according to the quantum theory its vibrational energy can only change by $h\nu_1$, so that the frequencies which will be emitted when the vibrational and rotational

energies both change will be given by

$$\nu = \nu_1 + \frac{h}{8\pi^2 K}\{n^2 - (n \pm 1)^2\},$$

or

$$\nu = \nu_1 + \nu_0(\mp 2n - 1).$$

In this case, when ν_1 is greater than ν_0, as is usually the case, n can change to $n + 1$ or to $n - 1$, with emission of light. The frequencies emitted are

n		ν	
4	....	$\nu_1 + 7\nu_0$	$\nu_1 - 9\nu_0$
3	....	$\nu_1 + 5\nu_0$	$\nu_1 - 7\nu_0$
2	....	$\nu_1 + 3\nu_0$	$\nu_1 - 5\nu_0$
1	....	$\nu_1 + \nu_0$	$\nu_1 - 3\nu_0$
0	....	$\nu_1 - \nu_0$	$\nu_1 - \nu_0$

Thus a band is obtained with frequencies $\nu_1 + \nu_0$, $\nu_1 - \nu_0$, $\nu_1 + 3\nu_0$, $\nu_1 - 3\nu_0$, and so on, but not with frequency ν_1. Such bands are observed in the infra-red spectra of HCl, HBr, and other diatomic molecules. The line $\nu_1 + \nu_0$ is usually absent.

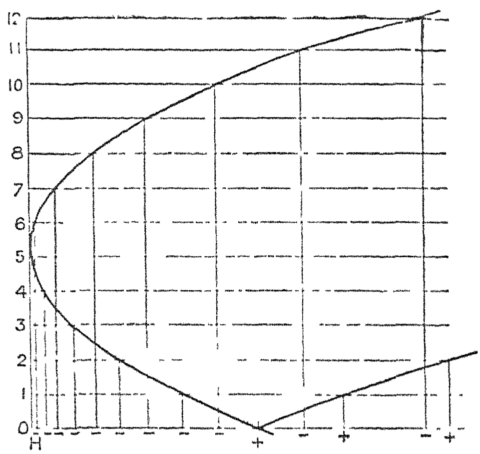


Fig. 2

So far we have supposed that the moment of inertia of the molecule is a constant independent of its rotational and vibrational energies, and that the internal energy of the molecule does not change. In the case of bands in the visible or ultra-violet regions it is necessary to suppose that the internal energy of the molecule changes as well as the rotational and vibrational energies. The change in the internal energy, however, may be supposed to involve a change in the moment of inertia and so of ν_0. The frequencies emitted are therefore given by

$$\nu = \nu_1 + \nu_2 + \nu_0 n^2 - \nu_0'(n \pm 1)^2,$$

where ν_0' is the new value of ν_0, and ν_2 corresponds to the change in the internal energy.

This gives $$\nu = A + Bn + Cn^2,$$

where $$A = \nu_1 + \nu_2 - \nu_0', \quad B = 2\nu_0', \quad C = \nu_0 - \nu_0'.$$

It is found that many bands can be represented very well by this expression. The frequencies given by $\nu_- = A - Bn + Cn^2$ are said to form the negative branch of the band, and those given by $\nu_+ = A + Bn + Cn^2$ the positive branch.

The head of the band is at the minimum value of ν on the negative branch, and is given by $\frac{d\nu}{dn} = -B + 2Cn = 0$, or $n = \frac{B}{2C}$, so that

$$\nu = A - B\frac{B}{2C} + C\frac{B^2}{4C^2} = A - \frac{B^2}{4C}.$$

If $B/2C$ is not equal to an integer then the head is at the value of ν given by the integer nearest to $B/2C$. The relation between ν and n can be made very clear graphically, as shown in fig. 2.

The curves represent n as a function of ν, and horizontal lines are drawn to represent the positive integral values of n. The values of ν in the band are then indicated by the intersections of the horizontal lines and the curves. The head of the band is at H.

REFERENCES

1. *Report on Series in Line Spectra.* A. Fowler.
2. *Atomic Structure and Spectral Lines.* A. Sommerfeld.
3. *The Theory of Spectra and Atomic Constitution.* N. Bohr.

CHAPTER IX

Cathode Rays, β-Rays, and α-Rays

CATHODE RAYS

1. Crookes' Experiments.

Cathode rays were discovered by Plücker in 1859; he observed that when an electric discharge was passed between two electrodes through a tube and the gas pressure was reduced to a sufficiently small value, the glass walls of the tube near the cathode emitted a greenish-coloured light. This light could be moved about by bringing a magnet near the tube, and appeared to be produced by something coming from the cathode and striking the glass. Cathode rays were later investigated by Hittorf and Goldstein and in 1879 by Crookes, whose beautiful experiments made clear many of the properties of the rays in a striking way.

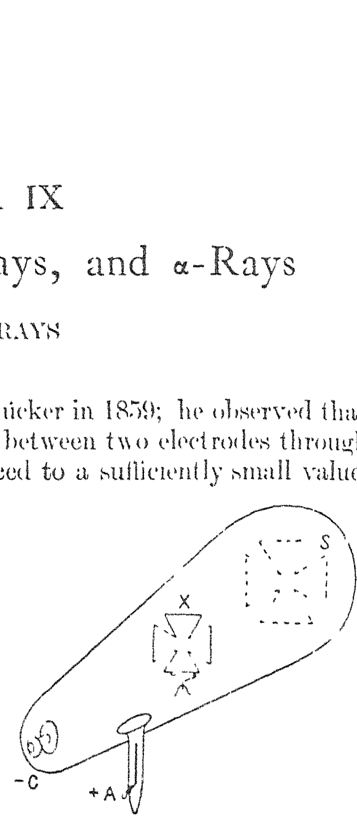


Fig. 1

As the pressure in a discharge tube is reduced the distance between the negative glow and the cathode increases, and when the Crookes dark space reaches the walls of the tube the green light observed by Plücker becomes very bright.

If a solid body is put up between the cathode and the walls of the tube a sharply defined shadow of it appears in the green light emitted by the walls. This was clearly shown by Crookes with the famous "Crookes tube" shown in fig. 1. The glass tube is conical in shape with a nearly flat end at S. At C a small aluminium disc supported by a wire sealed through the glass serves as the cathode. The anode is a wire A in a side tube. At X a piece of mica or aluminium sheet in the form of a cross is supported on a hinge so that by tilting the tube it can be put in either a horizontal or vertical position. The gas pressure in the tube is reduced to about 0·01 mm., so that the Crookes dark space fills the tube and the greenish light is emitted by the flat

end when the cathode and anode are connected to an induction coil. When the cross is up it throws a sharp shadow on the end of the tube, showing clearly that something is emitted by the cathode which is stopped by the cross. Also the shadow is sharp although the source is of considerable size, showing that the rays are not emitted in all directions from every point on the cathode but in only one direction It is found that the rays are emitted in a direction perpendicular to the surface of the cathode close to its suface. Thus Crookes showed that by using a concave cathode the rays could be focused on to a small area, and if a piece of platinum foil was put up at the focus the foil became very hot and could even be melted. The cathode rays are deflected by a magnetic field as we should expect negatively charged particles to be deflected. A Crookes tube for showing this is shown in fig 2.

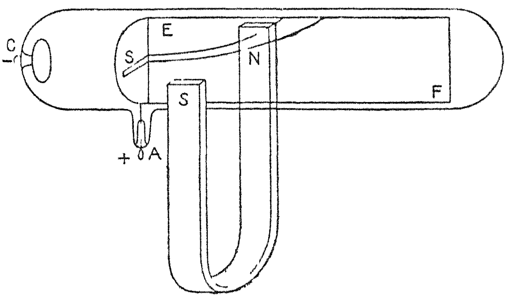


Fig 2

The cathode is a disc at C. Parallel to it a few centimetres away is a metal sheet, with a slit S cut in it, which forms the anode. The narrow beam of cathode rays which pass through the slit falls on a slightly inclined screen EF coated with calcium sulphide, or some other substance, which fluoresces brightly when struck by the rays. A bright streak of light appears on the screen which can be deflected by a magnetic field perpendicular to the screen. In the figure a magnet is shown with its S pole in front and its N pole behind the tube, deflecting the rays upwards as shown.

2. The Wehnelt Cathode. Magnetic and Electric Deflection.

It was found by Wehnelt that calcium and barium oxides emit cathode rays very freely when heated to a red heat and negatively charged By means of a cathode consisting of a strip of platinum foil, which can be heated by a current, having a small area on it coated with calcium or barium oxide an intense narrow beam of cathode rays can be obtained. If the gas pressure is not too low the narrow beam causes the gas to emit light so that its path through the tube can be easily seen in a dark room and the deflection of such a beam by a magnetic field observed. A tube with a Wehnelt cathode is shown in fig. 3. The anode A is a small aluminium disc and the cathode C a strip of platinum foil about 2 mm. wide supported by two wires E and F. A small patch of oxide is on the strip near its middle. E and F are

connected to a battery and the strip heated to a dull red heat. A potential difference of a few hundred volts between C and A is then sufficient to produce an intense beam of cathode rays from the oxide patch which comes out perpendicular to the surface of the foil and diverges slightly. By means of a magnet the beam of rays can be deflected along a curve as shown in the figure.

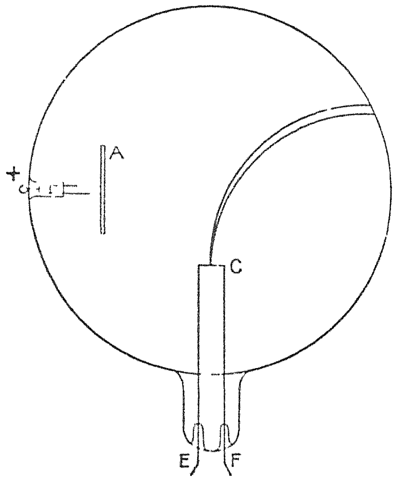


Fig 3

The potential difference required to produce a discharge in a Crookes tube, in which the dark space fills the tube is usually of the order of 10,000 volts or more, which is much greater than is necessary with a Wehnelt cathode. The magnetic deflection of cathode rays increases as the potential difference used to produce them decreases, so that the rays from a Wehnelt cathode using a few hundred volts are much more easily deflected than the rays in a Crookes tube. Cathode rays are deflected by an electric field like negatively charged particles. A tube used by J J. Thomson to show this is shown in fig. 1. The cathode is a small aluminium disc at C. The anode is a metal disc A having a small hole in it through which a narrow beam of the rays from C passes. The beam is further limited by a second disc B with a small hole in it, and then passes between two parallel metal plates D and E to the end of the tube near P, where it produces a luminous spot on the glass If D and E are connected together the rays pass along the axis of the tube, but if a potential difference is maintained between them by means of a battery the rays are deflected towards the positively charged plate.

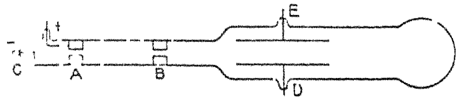


Fig 1

That the cathode rays carry a negative charge was first shown directly by Perrin in 1895. He passed a beam of the rays into a metal cylinder and found that the cylinder received a negative charge. Hertz noticed that cathode rays can pass through very thin metal foil, and Lenard showed that they can be passed through a thin metal window in a discharge tube into the air outside or into another tube containing a gas at any desired pressure.

3. Cathode Rays are Negatively Charged Particles. Charge and Mass.

The experimental results described above show that cathode rays are negatively charged particles moving with a high velocity, and for a time it was generally believed that they were negatively charged atoms of the gas in the discharge tube or of the metal of the cathode. About 1897, however, several physicists succeeded in measuring the ratio of the charge e to the mass m of the cathode ray particles and found this ratio to have a surprisingly high value, about 10^7 electromagnetic units per gramme. It was also shown by Lenard that they have considerable penetrating power and can pass through several millimetres of air at atmospheric pressure.

Before 1897 the highest ratio of charge to mass known was that for the positively charged hydrogen ion in solution for which e/m is about 10^4 E.M. units per gramme. The charge on a univalent ion in solutions was believed to be a sort of atomic unit of electricity because ionic charges in solutions are always multiples of it, so that physicists were inclined to expect that gaseous ions would be found to carry the same charges as ions in solutions. The high value of e/m for cathode rays might have been explained by supposing e to be about 1000 times the charge on one hydrogen ion in solutions and m equal to the mass of one hydrogen atom. J. J. Thomson, who was one of the first to measure e/m for cathode rays, pointed out that the cathode rays could not have the high penetrating power observed by Lenard if they were of atomic dimensions, and suggested that the high value of e/m was due to m being about 1000 times smaller than the mass of one hydrogen atom, the charge e being the same as that on one univalent ion in solutions. J. J. Thomson supported this revolutionary theory by many ingenious experiments, and it is now universally accepted. Other physicists, notably Wiechert and Kaufmann, also measured e/m for cathode rays at the same time as J. J. Thomson, and Wiechert suggested the same explanation of its high value, but to J. J. Thomson as the principal advocate of the new theory belongs the greater share of the credit for it.

The cathode ray particles or electrons, as they are now called, can be obtained from all kinds of matter, and are one of the few constituents of which atoms appear to be composed. The discovery of particles 1000 times lighter than hydrogen atoms, obtainable from all kinds of matter, may be said to mark the beginning of modern physics as distinguished from the classical physics of Faraday, Clerk-Maxwell, and Helmholtz.

4. Ratio of Charge to Mass. Measurements of Kaufmann and J. J. Thomson.

The methods used to measure e/m for cathode rays will now be considered. The apparatus used by Kaufmann is shown in fig. 5. The cathode C and anode A were connected to a Wimshurst machine, and the potential difference V between them was measured with an electrostatic voltmeter. The anode was a thin straight wire, and the cathode rays produced a shadow of it on the plate P which closed the lower end of the tube. A nearly uniform magnetic field perpendicular to the rays and parallel to the anode wire was produced by a large coil SS′ which surrounded the part of the tube below the anode as shown. The magnetic field deflected the rays and the deflection of the shadow of the anode on the plate P was measured. The velocity v of an electron starting from the cathode is given by the equation

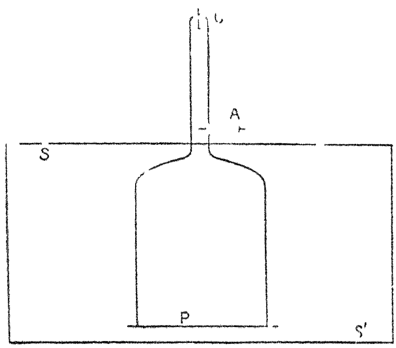


FIG. 5

$$Ve = \tfrac{1}{2}mv^2,$$

where e is the charge and m the mass of the electron. In the magnetic field (H) the electrons move along a circular path of radius r given by $mv^2/r = Hev$. The radius r was calculated from the observed deflection and the dimensions of the apparatus. These equations give $e/m = 2V/H^2r^2$, and $v = 2V/Hr$. The potential difference V was varied from about 3000 to 10,000 volts by reducing the gas pressure in the tube from 0·07 mm to 0·03 mm. The final result obtained was

$$\frac{e}{m} = 1{\cdot}77 \times 10^7 \text{ E.M. units per gramme.}$$

A possible objection to this method is that some of the electrons may be liberated in the gas between the cathode and anode and so not fall through the whole of the potential difference V. Such electrons, however, would not produce a sharp shadow on the screen because they would be deflected through different distances depending on their velocity. The sharp shadow which was obtained must have been due to electrons all having the same velocity, so that they must have started at the surface of the cathode as Kaufmann assumed. In discharges at low pressures the cathode rays are mainly produced by the impact of positive ions on the cathode.

J. J. Thomson, using a tube like that shown in fig. 1, measured the deflection of the rays by a magnetic field and then balanced the

magnetic deflection by the electric field between the plates D and E. The magnetic field was perpendicular to the rays and parallel to the surfaces of the plates D and E, and extended over the space between these plates. In this case $Fe = Hev$, where F is the electric field strength, so that $v = F/H$. For the magnetic deflection $mv^2/r = Hev$, so that $e/m = F/H^2r$. In this way J. J. Thomson found e/m to be about 10^7, and v to be about 3×10^9 cm. per second when the gas pressure was 0·01 mm.

J. J. Thomson also measured e/m by another method. A beam of the rays was received in a small metal cylinder, and the rate of rise of temperature of this cylinder was determined with a thermocouple. The charge received by the cylinder was measured by connecting it to a galvanometer. In this way the energy W and charge Q of the rays were found. We have then $W/Q = \frac{1}{2}mv^2/e$, so that if v is found from the magnetic and electric deflections or otherwise, e/m can be calculated. In this way J. J. Thomson found e/m to be about 10^7, as in his previous experiments.

The value of e/m for cathode rays has, since 1897, been carefully measured by many physicists. The value now accepted for slowly moving electrons is $1{\cdot}760 \times 10^7$ E.M. units per gramme. It was found that e/m for the cathode rays is the same for discharges in different gases and with cathodes made of different metals. It is also the same for the cathode rays emitted by hot bodies or set free by ultra-violet light, and for the β-rays or electrons emitted by radioactive bodies. It appears that electrons of identical properties can be obtained from any kind of matter; they must therefore be one of the constituents of the chemical atoms.

5. Cathode Rays and Ionization of Gases.

When cathode rays are passed through a gas they produce ions by collisions with the gas molecules. The ionization produced by electrons moving through a gas under the action of a uniform electric field is discussed in the chapter on the motion of electrons in gases. When cathode rays are passed into a gas the velocity with which they move is gradually diminished by collisions with the gas molecules. High velocity rays travel in straight lines through gases at low pressures with little loss of energy, but slower rays are deflected by the molecules and describe paths of irregular shape until they are stopped. The number of pairs of ions produced by cathode rays per centimetre of path has been determined by Durack and by Glasson. Durack found that cathode rays produced 0·4 pair of ions and that β-rays from radium produced 0·17 pair of ions per centimetre in air at 1 mm. pressure. The velocity of the cathode rays was probably about 4×10^9 cm. per second and that of the β-rays about 2×10^{10} to $2{\cdot}8 \times 10^{10}$ cm. per second.

Glasson obtained cathode rays of known velocity by deflecting them by a uniform magnetic field and obtained the following results:

Velocity in Cm./sec.		Pairs of Ions produced per Cm. in Air at 1 Mm. Pressure
4.08×10^9	. . .	2.01
4.76×10^9	. .	1.53
5.44×10^9	. . .	1.26
6.12×10^9	. . .	0.99

It appears that the ionization produced diminishes as the velocity increases. The number of molecules struck by an electron in going 1 cm. through air at 1 mm. pressure must, according to the kinetic theory of gases, be about 180. For more slowly moving electrons Townsend found the maximum number of pairs of ions formed to be 11·6 per centimetre in air at 1 mm. pressure.

6. Absorption of Cathode Rays.

The absorption of cathode rays by thin metal sheets and by gases was investigated by Lenard in 1895. The cathode rays in a discharge tube were passed, through a small window of aluminium foil in the wall of the tube, into another tube. A screen of barium platinocyanide was put up in the second tube opposite the window so that the cathode rays fell on it and caused it to phosphoresce. The intensity of the rays at the screen was estimated by the brightness of the luminosity of the screen. The absorption of the rays by thin sheets of metal and by gases at different pressures was determined by interposing them between the window and the screen and measuring the diminution of the luminosity. It was found that the intensity of the rays transmitted was nearly proportional to $\varepsilon^{-\lambda t}$ where λ is a constant and t the thickness of the absorbing substance. The constant λ depends on the nature of the absorbing substance and on the velocity of the rays. It was found that for rays produced by a potential difference of 30,000 volts the constant λ was nearly proportional to the density d of the absorbing layer. The following table gives some of Lenard's results.

Substance.		λ cm.$^{-1}$	d gm./cm.^{3}	λ/d
Hydrogen at 3 mm. pressure	. .	0·00149	3.5×10^{-7}	4,040
Hydrogen at 760 mm. pressure	. .	0·476	8.5×10^{-5}	5,610
Air at 760 mm. pressure	. .	3.42	1.2×10^{-3}	2,780
Paper	. .	2,690	1.3	2,070
Mica	. .	7,250	2.8	2,590
Gold	. .	55,600	19.3	2,880

Thus it appears that for a range of d from 3.6×10^{-7} to 19·3 the quotient λ/d only varies by a factor of less than 3. This extremely interesting result was obtained by Lenard in 1895. If M denotes the mass per unit area of the absorbing layer then $M = dt$, so that $\lambda t = \lambda M/d$. Since λ/d is nearly the same for all substances this shows that the absorption depends almost entirely on the mass of the screen and not on the kind of matter of which it is made. According to the modern theory of atoms each atom consists of a positively charged nucleus sur-

rounded by a number of electrons. The number of electrons is roughly half the atomic weight, so that the number of electrons per cubic centimetre in any kind of matter is roughly proportional to the density of the matter. In the case of heavy atoms like gold the number of electrons is much greater than the number of nuclei, but in the case of hydrogen there is only one electron to each nucleus. Thus the sum of the numbers of electrons and nuclei per unit mass of matter for hydrogen is about double its value for the heavier elements. The absorption of the cathode rays may be attributed to collisions with the electrons and nuclei, so that we should expect the ratio λ/d for hydrogen to be about double its value for the heavier elements, as was found by Lenard to be the case. The absorption of cathode rays is nearly the same for different kinds of matter because all kinds of matter consist of electrons and nuclei. The electrons and nuclei only occupy a very minute fraction of the volume of the matter so that cathode rays are able to pass through. We should expect any kind of atom to pass through matter as easily as cathode rays, but we cannot obtain atoms moving with the high velocity of cathode rays.

W. Wilson measured the velocity V of cathode rays before and after passing them through thin sheets of metal. The velocity was found from the radius of the circle described in a magnetic field of known strength. He found that

$$V_0^4 - V^4 = kt,$$

where V_0 is the velocity before entering the sheet, V that after passing through, t the thickness of the sheet, and k a constant. The rays are deflected in the sheet by collisions with the atoms, so that the distance they actually travel in the sheet may be considerably greater than t. Since the energy of the rays is nearly proportional to V^2 and the number of collisions to t, this indicates that the energy lost per collision is inversely as the energy of the rays.

β-RAYS

7. Kaufmann's Experiments.

The β-rays which are emitted by many radioactive substances are found to be of precisely the same nature as cathode rays. They carry a negative charge and are deflected by electric and magnetic fields. The velocities of β-rays range up almost to the velocity of light.

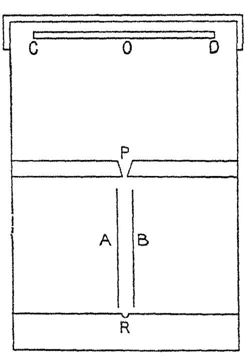


Fig 6

The ratio e/m for β-rays was investigated accurately by Kaufmann and later by Bucherer. Kaufmann's apparatus is shown in fig. 6. A particle of radium bromide was placed at R in a depression in a thick metal plate. The β-rays from the radium passed between two parallel metal plates A and B to a small hole P in a thick metal sheet. From P the narrow beam of rays passed on to a photographic plate CD. A potential difference was maintained between A and B by means of a battery, and the whole apparatus was placed in a uniform magnetic field parallel to the direction of the electric field between A and B. The magnetic field deflects the rays so that they move along circular

paths, in planes perpendicular to the field, of radius r given by $mv^2/r = Hev$. The circular path of the rays is shown in fig. 7. The radium is at R and the rays pass through the small hole at P and fall on the photographic plate at Q so that R, P, and Q lie on the circle. Let the magnetic deflection $OQ = y$, $RO = a$, and $RP = b$. Then we have

$$r^2 = \frac{b^2}{4} + \left(\frac{a^2 - ab + y^2}{2y}\right)^2,$$

so that r can be calculated when y, a, and b have been measured.

The electric deflection, which will be denoted by z, is perpendicular to y on the photographic plate. To calculate z approximately we may assume the electric field uniform, and equal to F from R to P, and to zero from P to the plate. The path of the rays in the plane of the electric field is shown in fig. 8. It is parabolic from R to P, and straight from P to Q′. The acceleration of the electrons due to the electric field is Fe/m, so that approximately

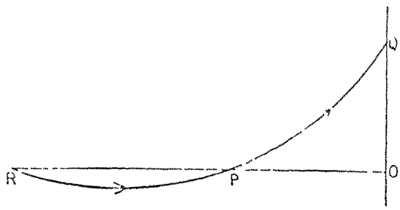


Fig. 7

$$\tfrac{1}{2}\left(\frac{Fe}{m}\right)\left(\frac{b}{v}\right)^2 - v'\frac{b}{v} = 0,$$

where v' is the initial transverse velocity component of the rays which get to P. Hence $v' = \frac{1}{2}\left(\frac{Fe}{m}\right)\frac{b}{v}$. The transverse velocity at P due to

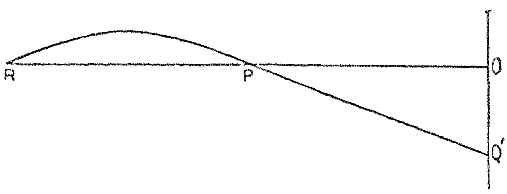


Fig. 8

the electric field is equal to $\frac{Fe}{m}\frac{b}{v}$, so that the resultant transverse velocity at P is $\frac{1}{2}\frac{Fe}{m}\frac{b}{v}$. Hence

$$z = (a - b)\frac{\frac{1}{2}\frac{Fe}{m}\frac{b}{v}}{v},$$

or

$$z = (a - b)\frac{b}{2}\frac{Fe}{mv^2}.$$

This equation and $mv/r = He$ enable v and e/m to be calculated. The β-rays emitted by radium bromide have velocities varying from 10^{10} cm. per second to almost 3×10^{10} cm. per second. Kaufmann thus obtained a curve on his photographic plates the co-ordinates of points on which were y and z. Each point on the curve corresponds to rays having a definite velocity, so that by measuring y and z for different points on the curve e/m could be found for rays having different velocities. Kaufmann found that e/m decreases as v increases. The variation observed agrees nearly with the equation

$$\frac{e}{m} = \frac{e}{m_0}\sqrt{1 - v^2/c^2},$$

where c is the velocity of light and e/m_0 is the value of e/m when v/c is small.

8. **Bucherer's Experiments. Mass and Velocity.**

The apparatus used by Bucherer is shown in fig. 9. AB and CD are two parallel circular metal discs only a fraction of a millimetre apart. A small particle of radium, R, was put between them at their centres. The β-rays from the radium moved out from R radially in all directions between the discs and fell on a cylindrical photographic film PP′ which was concentric with the discs. The discs and film were in a metal box in which a good vacuum was maintained. An electric field F was produced between the discs by connecting them to a battery, and the whole apparatus was put in a uniform magnetic field H parallel to the plane of the discs. Since the discs were so near together the electrons from the radium could not get to the photographic film unless the force on them due to the electric field was equal and opposite to that due to the magnetic field. We have therefore, for rays going along a radius making an angle θ with the magnetic field,

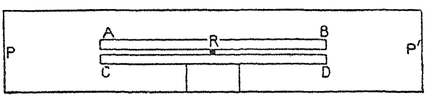


Fig. 9

$$Fe = Hev \sin\theta, \quad \text{or} \quad v = F/H \sin\theta.$$

Bucherer made F/H equal to $\frac{1}{2}c$ in most of his experiments, so that $v/c = 1/2 \sin\theta$. If $v = c$ this gives $\sin\theta = \frac{1}{2}$, so that $\theta = 30°$ or $150°$. Now v is never greater than c, so that no rays should get to the film except between $\theta = 30°$ and $\theta = 150°$. Bucherer found that this was the case. When $\theta = \pi/2$ then $v = \frac{1}{2}c$, which is therefore the smallest velocity of the rays which can get to the film when F/H is equal to $\frac{1}{2}c$. The rays which get out from between the discs are deflected by the magnetic field and so give a curve on the photographic film. The

deflection is approximately equal to $\frac{1}{2}a^2He\sin\theta/mv$, where a is the difference between the radius of the film and that of the discs, for the magnetic field gives the electrons an acceleration $Hev\sin\theta/m$ which acts for a time a/v.

In this way Bucherer found v and e/m for several values of v. The following table gives his results.

v/c	$\dfrac{e \times 10^{-7}}{m\sqrt{1-v^2/c^2}}$
0·3173	1·752
0·3787	1·761
0·4281	1·760
0·5154	1·763
0·6870	1·767

The second column gives the observed values of e/m multiplied by 10^{-7} and divided by $\sqrt{1-v^2/c^2}$. It appears that this quantity is practically constant and equal to $10^{-7} \times e/m$ for cathode rays, which is 1·769. If we assume that the charge e is independent of v then these experiments show that

$$m = \frac{m_0}{\sqrt{1-v^2/c^2}},$$

where m_0 is the mass when $v = 0$.

This variation of the mass with the velocity is shown in the chapters on the electron theory and on relativity to be due to energy possessing mass equal to the energy divided by the square of the velocity of light, so that when the kinetic energy of a particle of any kind is increased then its mass is also increased. At one time it was supposed that the observed variation of the mass of electrons with their velocity gave information as to the nature of these particles, but it is now clear that if energy possesses momentum equal to Ev/c^2, where E is the energy, v its velocity, and c the velocity of light, then the mass of any particle varies with its velocity so that

$$m = \frac{m_0}{\sqrt{1-v^2/c^2}}.$$

The observed variation of the mass of electrons with their velocity is therefore presumably the same as for all other bodies and gives no information about their constitution or the distribution of the charge in them. So far as we know, electrons behave like particles of mass m and charge e, so that the equation of motion of an electron is

$$\frac{d}{dt}(mv) = F,$$

where F is the resultant force on the electron and mv its momentum.

9 Scattering of β-rays by Matter. Mathematical Theory.

When β-rays are passed through thin sheets of matter they are scattered like cathode rays, but owing to their higher velocities the scattering is less. If a narrow parallel beam of β-rays is allowed to fall normally on a thin metal sheet, of thickness t, then the rays which emerge on the other side form a diverging beam the intensity in which is greatest along the normal to the sheet and falls off gradually as the inclination to the normal is increased. The angle between the emerging β-rays and the normal, or the deviation from their original direction, is called the scattering angle and will be denoted by ϕ. In passing through the sheet any particular β-ray passes through a large number of atoms, so that the angle ϕ through which it is finally deviated is the resultant of a large number of small deviations.

Let the initial direction of the beam be horizontal and let the deviation of a β-ray in the horizontal plane be x and in the vertical plane y, so that when ϕ is small we have $\phi^2 = x^2 + y^2$. Let the number of rays for which x is between x and $x + dx$ be

$$n_0 f(x)\,dx,$$

where n_0 is the total number of β-rays considered, and $f(x)$ denotes a function of x only. In the same way let $n_0 f(y)dy$ denote the number for which y is between y and $y + dy$. Then, since it is clear that the distribution as regards x is independent of that as regards y, the number for which x is between x and $x + dx$ and also y between y and $y + dy$ must be

$$n_0 f(x) f(y)\,dx\,dy.$$

But this must be a function of $x^2 + y^2$ since the distribution must clearly be symmetrical about the origin at $x = 0$ and $y = 0$. Hence

$$n_0 f(x) f(y)\,dx\,dy = n_0 \psi(x^2 + y^2)\,dx\,dy,$$

where $\psi(x^2 + y^2)$ denotes some function of $x^2 + y^2$. Hence

$$f(x) f(y) = \psi(x^2 + y^2).$$

Differentiating this with respect to x and dividing one side by $f(x)f(y)$ and the other by $\psi(x^2 + y^2)$ we get

$$\frac{f'(x)}{f(x)} = 2x\frac{\psi'(x^2 + y^2)}{\psi(x^2 + y^2)}.$$

In the same way

$$\frac{f'(y)}{f(y)} = 2y\frac{\psi'(x^2 + y^2)}{\psi(x^2 + y^2)},$$

so that

$$\frac{1}{2x}\frac{f'(x)}{f(x)} = \frac{1}{2y}\frac{f'(y)}{f(y)}.$$

Hence both these quantities must be equal to a constant, say $-\alpha$, so that

$$\frac{f'(x)}{f(x)} = -2\alpha x,$$

which gives

$$\log f(x) = -\alpha x^2 + \text{constant},$$

or

$$f(x) = A\varepsilon^{-\alpha x^2},$$

where A is another constant. To determine the constants we have

$$n_0\int_{-\infty}^{+\infty} f(x)\,dx = n_0,$$

and, if $\bar{x}^2$ denotes the average value of x^2,

$$n_0\int_{-\infty}^{+\infty} x^2 f(x)\,dx = n_0\bar{x}^2.$$

These equations give $A\sqrt{\frac{\pi}{\alpha}} = 1$ and $\frac{A}{2\alpha}\sqrt{\frac{\pi}{\alpha}} = \bar{x}^2$, so that $\alpha = \frac{1}{2\bar{x}^2}$. Hence

$$f(x) = A\varepsilon^{-x^2/2\bar{x}^2},$$

and in the same way $$f(y) = A\varepsilon^{-y^2/2\bar{y}^2}.$$

Also, if $\overline{\varphi^2}$ denotes the mean value of φ^2 then, since $\varphi^2 = x^2 + y^2$, we have $\overline{\varphi^2} = \overline{x^2} + \overline{y^2} = 2\overline{x^2}$, for $\overline{x^2}$ must be equal to $\overline{y^2}$. Hence

$$\psi(\varphi^2) = f(x)f(y) = A^2\varepsilon^{-\frac{x^2+y^2}{2\overline{x^2}}} = A^2\varepsilon^{-\phi^2/\overline{\phi^2}}.$$

The number of β-rays for which φ is between φ and $\varphi + d\varphi$ is therefore

$$n_0A^2\varepsilon^{-\phi^2/\overline{\phi^2}}2\pi\varphi d\varphi,$$

or $$\pi n_0A^2\varepsilon^{-\phi^2/\overline{\phi^2}}d(\varphi^2).$$

The number for which φ is less than a given value φ is therefore

$$\pi n_0A^2\int_0^{\phi}\varepsilon^{-\phi^2/\overline{\phi^2}}d(\varphi^2) = \pi n_0A^2\overline{\varphi^2}(1 - \varepsilon^{-\phi^2/\overline{\phi^2}}).$$

Hence, if n_0 denotes the number of rays passing through the sheet and n the number for which the deviation is less than φ, we have

$$n/n_0 = 1 - \varepsilon^{-\phi^2/\overline{\phi^2}}.$$

The deviation φ is the resultant of the deviations due to individual atoms in the sheet Let x_1 denote the horizontal deviation due to a single atom and y_1 the vertical deviation. Then we have

$$x = \Sigma x_1,$$
$$y = \Sigma y_1,$$

where $x^2 + y^2 = \varphi^2$, and Σx_1 denotes the sum of all the horizontal deviations of an electron while passing through the sheet. Hence

$$x^2 = (\Sigma x_1)^2 = \Sigma x_1^2,$$

because the products of two different x_1's are as likely to be positive as to be negative, and so their sum must be zero. In the same way

$$y^2 = \Sigma y_1^2,$$

so that $$\varphi^2 = \Sigma(x_1^2 + y_1^2) = \Sigma\varphi_1^2,$$

where φ_1 is the deviation due to one atom. The mean value of φ^2 is therefore given by $$\overline{\varphi^2} = N\overline{\varphi_1^2},$$

where N is the number of atoms which a β-ray passes through in the sheet and $\overline{\varphi_1^2}$ is the mean of the squares of the atomic deviations. Hence we have

$$n/n_0 = 1 - \varepsilon^{-\phi^2/N\overline{\phi_1^2}}.$$

The number N will be approximately given by $N = \pi \mathcal{N} R^2 t$, where $\mathcal{N}$ is the number of atoms per cubic centimetre in the sheet and R the radius of an atom, for this is the number of atoms in a cylinder of radius R and length t which a β-ray going straight through the sheet would pass through. Hence

$$n/n_0 = 1 - \epsilon^{-\phi^2/(\pi \mathcal{N} R^2 \overline{\phi_1^2} t)}.$$

10. Experiments on Scattering of β-rays.

An apparatus used by Crowther to investigate the scattering of β-rays is shown in fig. 10. The radioactive body was placed at R in a vertical hole in a lead block. The β-rays from R passed out downwards and were deviated sideways by means of a magnetic field perpendicular to the plane of the paper, so that they traversed a circular path RAB, the beam being limited by diaphragms so that only rays having a definite velocity got to P. At P the rays fell on a thin sheet of metal and emerged from it as a diverging beam. The fraction of the rays which passed through a circular aperture S in a metal screen was measured by passing them into an ionization chamber Q containing air or some other gas, and finding the conductivity produced. The ratio of the conductivity with the scattering plate at P to the conductivity without any scattering plate is approximately equal to n/n_0. The scattering angle for the rays passing through the aperture S is less than $\tan^{-1}(a/d)$, where a is the radius of the aperture and d its distance from the scattering plate.

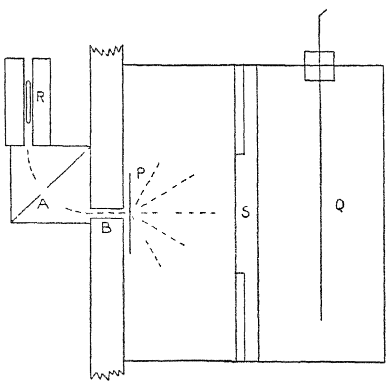


Fig 10

The results obtained by Crowther and other physicists agree approximately with the equation $n/n_0 = 1 - \epsilon^{-\phi^2/kt}$, where k is a constant. With different sized apertures the thickness of the plate required to produce a given value of n/n_0 is proportional to ϕ^2, in agreement with the theory. The following table gives some results obtained by Crowther and Schonland with $\phi = 0{\cdot}11$ for the scattering of β-rays by gold.

t	n/n_0 (Observed)	$t \log \frac{n_0}{n_0 - n}$	n/n_0 (Calculated)
1	0 86	0 85	0 89
2	0·67	0·96	0·67
3	0 52	0·96	0·52
4	0 46	1 07	0·43
5	0 38	1 03	0 36

The first column gives the number of gold foils used as the scattering plate, which is proportional to t. If $\frac{n}{n_0} = 1 - \epsilon^{-\phi^2/kt}$, then

$$\frac{\phi^2}{k} = t \log \frac{n_0}{n_0 - n},$$

so that the numbers in the third column should be constant according to the theory. The last column gives the calculated values of n/n_0, taking ϕ^2/k equal to 0·97, which is the mean of the numbers in the third column. Crowther and Schönland's results give $k = 680$ for gold. Now $k = \pi \mathcal{N} R^2 \overline{\phi_1^2}$, so that we can calculate $\overline{\phi_1^2}$ if we know $\mathcal{N}$ and R. The number of atoms of gold per cubic centimetre is 6×10^{22}, and we may assume $R = 10^{-8}$ cm., so that we get

$$\overline{\phi_1^2} = \frac{680}{\pi \times 6 \times 10^{22} \times 10^{-16}} = 3{\cdot}6 \times 10^{-5}.$$

According to this the root mean square deviation of the β-rays by an atom of gold was only 6×10^{-3} or about $\frac{1}{3}$ of a degree. The β-rays for which the above results were obtained had a velocity corresponding to a fall of potential of $4{\cdot}6 \times 10^5$ volts.

If we assume that the atoms consist of a positively charged nucleus surrounded by a number of electrons it is possible to calculate the average deviation of β-rays passing through such atoms, and the results obtained are consistent with the observed average deviations. For the details of such calculations the original papers may be consulted.

11. Absorption of β-rays.

When β-rays from a layer of radioactive material are absorbed by placing a thin sheet of metal over the layer, so that the rays fall on the sheet in all directions, then the intensity of the rays getting through is proportional approximately to $\varepsilon^{-\mu t}$, where μ is a constant and t the thickness of the sheet. The intensity may be measured by the ionization produced by the rays The constant μ is roughly speaking proportional to the density ρ of the material of the sheet, as with cathode rays. The following are some values of μ/ρ for the β-rays from uranium.

Substance	μ/ρ
Silver	7 3
Aluminium ..	4 1
Lead	9·75
Sulphur ..	4 5
NaCl . ..	4·7
KI . ..	7 8

For the β-rays from radium-E μ/ρ varies from 15·8 for carbon to 22·1 for tin. The following table gives, in centimetres, the thickness of aluminium penetrated by β-rays which describe a circle of radius R in a magnetic field of strength H perpendicular to the plane of the circle.

HR	=	1,380	2,535	3,790	5,026	7,490	11,370
Thickness	=	0·018	0·124	0·279	0·440	0·785	1·36

When β-rays pass through thin sheets of matter their velocity is decreased. The change in the product RH due to a sheet weighing 0 01 gm. per square centimetre is given approximately by

$$\Delta(RH) = K\frac{c^3}{v^3},$$

where v is the velocity of the rays, c that of light, $\Delta(RH)$ the decrease in RH and K a constant. The constant K is about 35 for mica, from 23 to 32 for tin, and about 28 for gold.

12. β-ray and Positron Energies.

The energies of the nuclear β-rays emitted by a number of radioactive elements have been measured by the magnetic focussing method

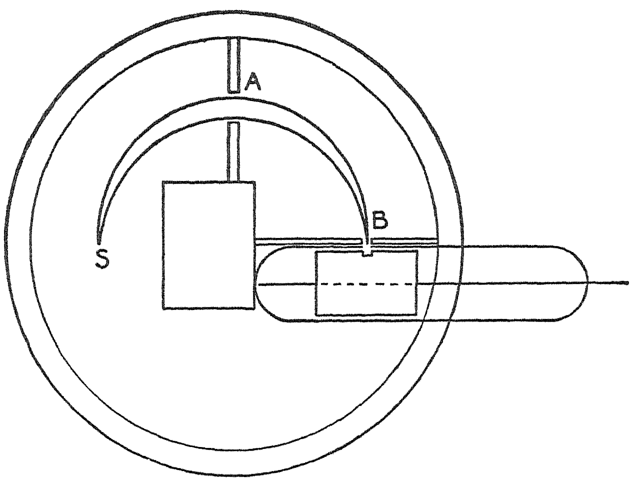


Fig. 11

used for the secondary β-rays and α-rays. The apparatus used by Scott is shown in fig. 11. The source of the rays S is a fine wire coated with the radioactive body. The rays from S describe circular paths in a magnetic field perpendicular to the plane of the diagram. Some of the rays pass through a wide slit at A and are focussed on a narrow slit at B. The rays passing through the slit at B enter a small Geiger counter through a thin cellophane window. The counter is connected to an amplifier and impulse counter which registers the number of β-rays. The energy of the rays can be calculated from the magnetic field strength H and the radius $\rho = \frac{1}{2}SB$ of the circular paths. We have $Hev = mv^2/\rho$, where v is the velocity, m the mass, and e the charge of the β-rays. Also $m = m_0/\sqrt{1 - v^2/c^2}$, where m_0 is the rest mass of the rays. The kinetic energy of the rays is equal to $c^2(m - m_0)$. If P denotes the potential difference in volts required

to give an electron energy equal to the β-ray energy, then these equations give

$$P = 299{\cdot}8\{((H\rho)^2 + (1703{\cdot}4)^2)^{1/2} - 1703{\cdot}4\},$$

where c has been put equal to $2{\cdot}998 \times 10^{10}$ cm./sec. and e/m_0 equal to $1{\cdot}760 \times 10^7$ e.m.u. per gramme. If we put $\tan\theta = H\rho/1703{\cdot}4$, then $P = 510682\ (\sec\theta - 1)$.

The number of β-rays counted varies with the strength of the magnetic field H. Fig. 12 shows the relation between the number per minute (n) and the field strength (H) for the β-rays from radium E. The path radius ρ was 1·98 cm. The number counted rises to a maximum as H is increased and falls to zero at a definite end point. Similar results have been obtained with other β-ray emitters. Some bodies emit two or more such groups of β-rays with different end points. Several artificially prepared radioactive elements emit positrons, and the positron energies may be found in the same way as the β-ray energies. The number of positrons counted varies in much the same way with the magnetic field strength as the number of β-rays.

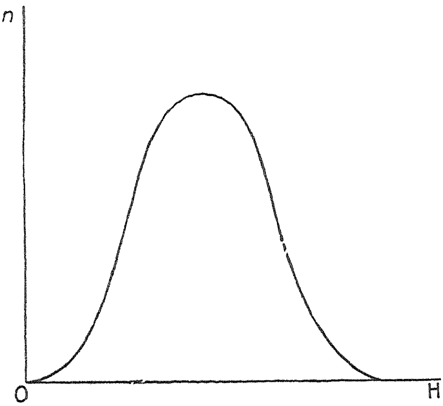


FIG 12

The following table gives the highest energies, corresponding to the end points, of the β-rays from several radioactive bodies.

Element	Rays	Highest Energy in Electron Volts.
RaE	β	$11{\cdot}7 \times 10^5$
RaC′	”	$31{\cdot}5 \times 10^5$
ThC′	”	$22{\cdot}5 \times 10^5$
${}_6C^{11}$	Positron	$11{\cdot}5 \times 10^5$
${}_{15}P^{30}$	”	37×10^5

The emission of β-rays and positrons with a continuous range of energies is not what might have been expected. Before the emission the atom appears to be in a definite state with a definite mass and energy, and after the emission also the new atom formed seems to have a definite energy. For example, RaC′ emits an α-ray with definite energy, the RaD formed emits a β-ray, and RaE formed next also emits a β-ray, and finally the RaF formed emits an α-ray always with the same definite energy. Thus we should expect the

sum of the energies of the two β-rays to be always the same. Actually the RaD β-rays all have very low energies, and the RaE β-rays have energies from zero to $11{\cdot}7 \times 10^5$ electron volts.

The heating effect due to the absorption of the β-rays from radium E has been measured and is found to be equal to that calculated from the energy distribution determined by the magnetic focussing method. This shows that the low-energy β-rays are not due to collisions between high-energy β-rays from the nucleus and extra-nuclear electrons. It seems very probable, therefore, that when an atomic nucleus disintegrates with the emission of a β-ray the energy of the β-ray may have any value between zero and a definite maximum value.

Radium-C, thorium-C and actinium-C disintegrate in two ways. They either emit an α-ray and then a β-ray, or first a β-ray and then an α-ray. It is found that the sum of the α-ray energy and the maximum β-ray energy has the same value for both ways of disintegration. The α-ray and maximum β-ray energies are different, but the sum of the two is the same. This suggests that when a β-ray is emitted the nucleus always loses energy equal to the maximum β-ray energy. When the energy of the β-ray is less than the maximum value the nucleus must lose energy, equal to the difference, in some other way. Pauli suggested that another particle must be emitted along with the β-ray with energy equal to the difference between the maximum β-ray energy and the energy of the β-ray. This new particle, Pauli supposed, has no charge and very small or zero rest mass, so that it has no interaction with matter and cannot be detected. These hypothetical particles are called neutrinos*.

13. Fermi's β-ray Theory.

Fermi has proposed a theory of nuclear β-ray and positron emissions. Fermi's theory is based on the analogy between the emission of a photon by an excited atom, and the emission of a β-ray and a neutrino by an excited nucleus. It is supposed that a neutron in the nucleus changes into a proton when a β-ray is emitted, and that a proton changes into a neutron when a positron is emitted. The chance of such a change taking place is supposed to depend on the action of the de Broglie waves of the emitted particles on the nucleus in much the same way that the chance of an excited atom emitting a photon may be supposed to depend on the action of the electromagnetic waves of the emitted photon on the electronic system of the atom. Since the nature of the action of de Broglie waves on the

* The existence of neutrinos has been confirmed by measuring the kinetic energy of atoms due to the emission of a neutrino. The recoil momentum of the atom is found to be equal and opposite to the calculated momentum of the neutrino.

neutrons and protons in a nucleus is unknown, Fermi assumed a hypothetical function to represent this action. His theory gives a relation between the mean life of the radioactive atoms and the maximum momentum of the β-rays or positrons. It also gives the distribution of energy for the emitted particles. The number N of radioactive atoms which have not yet disintegrated is given by $N = N_0 e^{-\alpha t}$, where N_0 is the number at time $t = 0$, so that the mean life is equal to $1/\alpha$. Fermi's relation may be written $\alpha^{-1}F(M) = \text{constant}$, where $F(M)$ denotes a function of M, the maximum momentum.

The following table gives some values of α^{-1} and $M = (H\rho)_{\text{Max}}/1703$, which is equal to the maximum momentum with m_0c as unit.

Element	Hours.	M.	$F(M)$.	$\alpha^{-1}F(M)$.
UX_2	0 026	5 4	115	3 0
RaB	0 64	2 04	1 34	0 9
ThB	15·3	1·37	0 176	2·7
ThC″	0 076	4·4	44	3 3
AcC″	0 115	3·6	17·6	2 0
RaC	0 47	7 07	398	190
RaE	173	3·23	10·5	1800
ThC	2 4	5 2	95	230
$MsTh_2$	8·8	6·13	73	640

It appears that the β-ray bodies fall into two groups. The first group has $\alpha^{-1}F(M)$ of order of magnitude unity, and the second order of magnitude about 100 times greater. For the first group the angular momentum of the β-rays about the nucleus is supposed to be zero, but for the second group it is equal to $nh/2\pi$ with n a small integer, according to Fermi's theory.

14. Neutrons.

Bothe and Becker, about 1930, discovered a very penetrating radiation emitted by beryllium and a few other light elements when bombarded with α-rays. This radiation was at first supposed to be γ-rays, but I. Curie and Joliot found that it caused the ejection of fast protons from paraffin, which γ-rays would not be expected to do. Chadwick investigated this penetrating radiation thoroughly and concluded that it consisted of high-velocity electrically neutral particles with masses about the same as protons. These new particles are called *neutrons* and are believed to be primary constituents of matter like electrons and protons.

When neutrons are passed through a C. T. R. Wilson cloud chamber containing hydrogen they produce no tracks themselves, but when a neutron collides with a hydrogen atom nucleus or proton, the proton makes a thick straight proton track quite different from a thin electron track. γ-rays also give no tracks, but they knock electrons out of atoms

which make thin tracks. Neutrons give no thin electron tracks. In nitrogen and other gases neutrons also produce short thick tracks due to collisions with the atoms but no neutron tracks. When the γ-rays from radiothorium which have energy $h\nu$ equal to 2·6 *MEV* are passed through heavy hydrogen or deuterium, they disintegrate the deuterium atoms into protons and neutrons thus:

$$h\nu + {}_1\mathrm{H}^2 \rightarrow {}_1\mathrm{H}^1 + {}_0n^1 + E.$$

The energy of the proton H^1 has been measured by finding its range and is found to be about 0·2 *MEV*. The neutron ${}_0n^1$ must have about the same energy as the proton, so that the energy required to disintegrate a deuterium nucleus is $2{\cdot}6 - 0{\cdot}4 = 2{\cdot}2$ *MEV*. This corresponds to 0·0024 atomic weight units. The atomic weight of ${}_1\mathrm{H}^2$ is 2·0147 and that of ${}_1\mathrm{H}^1$ is 1·0081, so that the atomic weight of the neutron is given by

$$2{\cdot}0147 = 1{\cdot}0081 + {}_0n^1 - 0{\cdot}0024,$$

which gives ${}_0n^1 = 1{\cdot}0090$. The neutron is therefore unstable since it can change into a proton with the emission of an electron with kinetic energy K given by

$$0{\cdot}00055 + 1{\cdot}0090 = 1{\cdot}0081 + 0{\cdot}00055 + K,$$

which gives $K = 0{\cdot}0009$ in atomic weight units or 0·84 *MEV*. It may be that 1·0090 is too high and that the atomic weight of the neutron is really less than the sum of the atomic weights of a proton and an electron or 1·0081. Neutrons penetrate matter easily because they carry no charge and so have no interaction with atoms except at exceedingly small distances. A beam of fast neutrons is reduced to one-half its intensity by about 5 gm./cm.^{2} of water or paraffin, or 50 gm./cm.^{2} of a heavy element like lead.

In substances containing hydrogen, like water and paraffin, fast neutrons are slowed down by collisions with the protons until they have velocities of the same order as those of gas molecules at the same temperature. Such slow neutrons are absorbed by most atomic nuclei, so producing a new atom with the same atomic number but with mass number greater by one unit. The new atoms so formed are often unstable and disintegrate immediately with the emission of a proton or α-ray. The element remaining after this emission is generally radioactive, emitting β-rays. This is discussed in the chapter on radioactivity.

α-RAYS

15. Ratio of Charge to Mass.

The properties of α-rays will now be considered. α-rays are rays emitted by many radioactive bodies; they are much less penetrating than the β-rays. They are deflected by electric and magnetic fields in the opposite direction to β and cathode rays, and are found to carry a positive charge.

The ratio of the charge e to the mass m of α-rays has been found by a method similar to that used by Bucherer for β-rays. The α-rays were passed between two parallel plates very near together and a potential difference was maintained between the plates. The apparatus was put in a uniform magnetic field perpendicular to the electric field between the plates and to the path of the rays. By properly adjusting the field strengths the force on the rays due to the electric field could be made equal and opposite to that due to the magnetic field, so that the rays were not deflected and could pass between the plates. In this case we have $Fe = Hev$, where F is the electric and H the magnetic field strength, so that the velocity v of the rays is given by $v = F/H$. After passing between the plates the rays emerge into the uniform magnetic field and describe circular paths in it of radius r given by $mv^2/r = Hev$. The rays fell on a photographic plate on which their deflection could be measured, and the radius r could be calculated from the deflection and the dimensions of the apparatus. In this way it was found that $e/m = 4823$ e.m.u. per gramme, and that v is between $1{\cdot}4 \times 10^9$ and $2{\cdot}2 \times 10^9$ cm. per second for the α-rays from different radioactive bodies.

16. Counting of α-rays. Scintillations. Ionization.

A remarkable property of α-rays was discovered by Sir William Crookes. He found that when α-rays fall on a screen coated with powdered zinc-blende the screen emits minute flashes of light easily visible in a low-power microscope. These flashes or scintillations are each due to the impact of one α-ray particle, so that when the rays are sufficiently feeble it is possible by counting the scintillations to determine the number of α-ray particles falling on the screen. Other fluorescent substances besides zinc-blende may be used—diamond, for example, makes a very good screen.

α-rays ionize gases strongly, and the ionization due to a single α-ray particle can be easily detected, which gives another method of counting the particles. The apparatus used by Rutherford to count the α-rays emitted by radium-C is shown in fig. 13. The radium-C was deposited on a small cone at S in a long glass tube EF provided with a stop-cock at T. At D there was a metal diaphragm having a small circular hole in it through which some of the α-rays from S

passed into a metal box in which there was an insulated electrode A. The potential difference between the box and the electrode, and the gas pressure in the apparatus, were adjusted so that an electron set free in the box produced a large number of ions by collisions with the

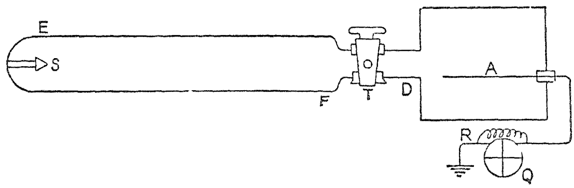


Fig 13

gas molecules In this way the ionization due to a single α-ray entering the box through D was made sufficient to cause an easily observable deflection of a quadrant electrometer connected to the insulated electrode A. The electrode A was also connected to the earth through a very high resistance R, so that the deflection due to an α-ray was not permanent but rapidly disappeared. The number of deflections due to α-rays was counted, and so the number passing through the hole at D in a known time found. The total number emitted by the radium-C could then be calculated by assuming the rays to be emitted equally in all directions.

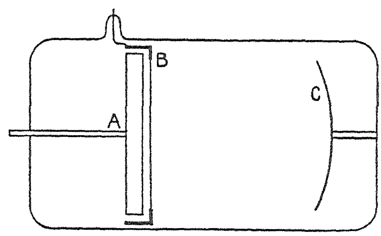


Fig 14

17. Charge carried by α-rays.

The charge carried by the α-rays emitted by radium-C was found with the apparatus shown in fig. 14. The radium-C was deposited on a plate C supported in a highly exhausted glass tube. The α-rays from C were received on an insulated electrode A after passing through a sheet of thin foil B and another sheet of thin foil covering the front of A. When α-rays strike a metal surface they cause the emission of electrons which leaves a positive charge on the metal in addition to the charge carried by the α-rays. To prevent these electrons from escaping from the electrode A it was covered with the thin foils which allowed the α-rays to pass through but stopped the electrons. The apparatus was also put in a strong magnetic field perpendicular to the plane of the paper which caused the electrons to describe small circular or spiral paths and so helped to prevent them from escaping. The charge received by the electrode A in a known time was measured with a quadrant electrometer. The total charge on the α-rays emitted by the radium-C in a known time could then be calculated by assuming

the rays to be emitted equally in all directions from every point on the plate at C. The amount of radium-C on the plate C was compared with that used in the previous experiment by comparing the intensities of the γ-rays emitted by the two deposits In this way it was found that an amount of radium-C equal to that contained in 1 gm of radium in radioactive equilibrium with its products emits $3{\cdot}4 \times 10^{10}$ α-rays per second and that the charge on these rays is 31 6 electrostatic units. The charge carried by one α-ray particle is therefore $9\,3 \times 10^{-10}$ electrostatic units. Since e/m is 4823 electromagnetic units per gramme it follows that the mass of one α-ray particle is given by

$$m = \frac{9{\cdot}3 \times 10^{-10}}{3 \times 10^{10} \times 4823} = 6{\cdot}42 \times 10^{-24} \text{ gm.}$$

The mass of one hydrogen atom is $1{\cdot}663 \times 10^{-24}$ gm., so it appears that the mass of one α-ray particle is nearly four times that of one hydrogen atom.

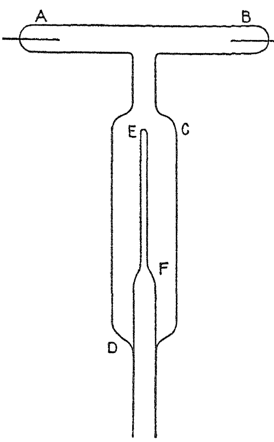


Fig. 15

18 **α-rays are Charged Helium Atoms.**

The ratio e/m for positively charged hydrogen atoms or hydrogen ions in solution is 9650, which is just double that found for α-rays. Also the charge $9{\cdot}3 \times 10^{-10}$ on one α-ray is nearly double the charge on one gaseous ion or electron. We conclude that the charge on one α-ray is $2e$ and that the mass of one α-ray particle is four times that of one hydrogen atom. Now the mass of one helium atom is known to be four times that of one hydrogen atom, which suggests that α-rays may be charged helium atoms.

That the α-rays are charged helium atoms was shown very clearly by an experiment due to Rutherford. The apparatus used is shown in fig. 15. A small discharge tube AB was connected to a tube CD and highly exhausted. The tube CD was sealed on to a tube EF, the end of which was drawn out into a narrow very thin walled tube. Some radium emanation was pumped into the narrow part of EF, and the α-rays emitted by the emanation passed through the thin glass walls into CD and AB. A discharge could then be passed through AB and the light emitted gave the spectrum of pure helium.

The helium atom is supposed to consist of a positively charged nucleus having a charge $2e$ and two electrons each having a charge $-e$. The α-ray particles are therefore supposed to be simply nuclei of helium atoms.

19. Range and Velocity of α-rays.

When α-rays from a very thin film of some radioactive body are passed through a gas it is found that they all travel very nearly the same distance before they are stopped. This distance is called the range of the α-rays. The range in a gas is found to be inversely as the gas pressure.

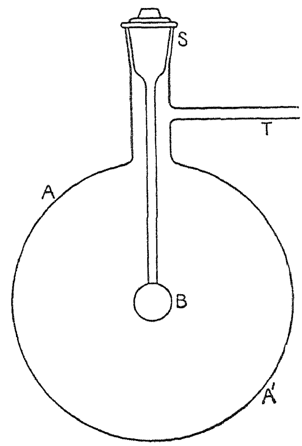


Fig. 16

The range may be determined by means of the apparatus shown in fig. 16. A spherical glass bulb AA′ is coated on the inside with a thin film of finely powdered zinc-blende, and the radioactive material which emits the α-rays is deposited on a small sphere B which is supported at the centre of the bulb on a rod attached to a stopper S. The gas pressure in the bulb can be varied by means of a side tube T which leads to a pump and manometer. The gas pressure is adjusted until the α-rays from B just reach the bulb, as shown by the scintillations produced on the zinc-blende film. The range is then equal to the radius of the bulb, and the range in the gas at 760 mm. pressure can be calculated since the range is inversely as the pressure. The range of the α-rays from radium-C is 7·0 cm. in air at 760 mm., and that of the rays from radium is 3·5 cm. The velocity of α-rays diminishes as they pass through a gas. Rutherford found the velocity of α-rays which had passed through different thicknesses of air by measuring the magnetic deflection of the rays, and found that

$$v^3 = A(R - x),$$

where v is the velocity of the rays, A a constant, R the range of the rays, and x the distance travelled from the source. Thus, when $x = R$, $v = 0$ as we should expect. Differentiating the above equation with respect to the time t we get

$$\frac{dv^2}{dt} = -\tfrac{2}{3}A,$$

so that it appears that the rate of loss of kinetic energy as the rays move through a gas is constant. Also $\dfrac{dv^2}{dx} = -\tfrac{2}{3}\dfrac{A}{v}$, so that the energy lost per unit distance moved is inversely as the velocity. We should expect the ionization produced to be proportional to the loss of energy, and therefore the ionization per centimetre to be inversely as $(R - x)^{1/3}$.

20. Ionization by α-rays. Stopping Power.

The ionization produced by α-rays along the range was investigated by Bragg with the apparatus shown in fig. 17 The α-rays were emitted by a film of radioactive material on a horizontal disc R, which carried a grating made of vertical metal plates so that only those α-rays moving in nearly vertical directions could get out. In this way a nearly parallel vertical beam of the rays was obtained. The rays passed through the gas in a box AB and through a horizontal plate of fine wire gauze at C Just above the gauze there was an insulated electrode D. The rays ionized the gas between C and D, and this ionization was measured by connecting D to a quadrant electrometer. Sufficient potential difference was maintained between C and D to saturate the current. By moving the source R up and down, the ionizations due to the rays at different distances from the source could be compared.

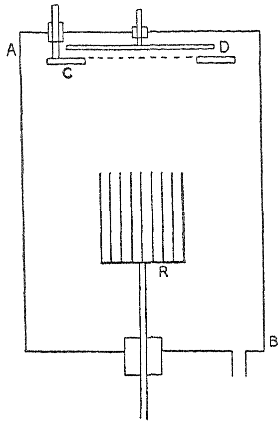


Fig 17

It was found that the ionization increased with the distance from the source to a maximum when the distance was nearly equal to the range of the rays and then rapidly fell to zero. The ionization was approximately proportional to $(R - x)^{-1/3}$ and therefore to the kinetic energy lost by the rays. The range in any gas is inversely as the pressure, so that the mass of gas per unit area in a layer of thickness equal to the range is independent of the gas pressure. This mass is $R\rho$ where ρ is the density of the gas. The *stopping power* of the gas is defined as $R\rho$ for air divided by $R\rho$ for the gas.

If α-rays fall normally on a very thin sheet of any solid and pass through, their velocity and range in air or other gas are diminished The decrease of the range may be found with the apparatus shown in fig. 17. If the thin sheet is put over the source R so that the rays pass through it, the decrease of the distance at which the maximum ionization is obtained is equal to the decrease of the range. The decrease of range due to a thin sheet is nearly proportional to the mass m per unit area of the sheet. If the weight of an atom of the material of the sheet is A, then m/A is the number of atoms per unit area in the sheet. Let ΔR denote the decrease of range in air of density ρ, and A' the mean weight of the air atoms. Then $\rho\Delta R/A'$ is the number of air atoms per unit area in a layer which diminishes the range by the same amount as the sheet. The *atomic stopping power* of the atoms in the sheet is defined as the ratio of the number of air atoms $\rho\Delta R/A'$ to the number

m/A of atoms per unit area in the sheet. The stopping power is therefore equal to $\rho A \cdot \Delta R/mA'$. The atomic stopping power of different elements is found to be roughly proportional to the square root of the atomic weight. For example, the atomic stopping power of silver is 3·1, and the square root of the ratio of the atomic weights of silver and air is $\sqrt{108/14 \cdot 4} = 2 \cdot 74$, which is not far from 3·1.

When α-rays go through thin metal sheets most of them go nearly straight through, but a small fraction is deviated through large angles. This scattering of α-rays has been investigated by Rutherford, Geiger, Chadwick, and others with very important results.

21. Single Scattering of α-rays.

The theory of the scattering of β-rays discussed earlier in this chapter was based on the assumption that the deviations were the resultants of a large number of small deviations, the average deviation due to passing through one atom being only a rather small fraction of a degree Such scattering is called *multiple scattering*. When α-rays are passed through matter the great majority are deviated through small angles, and these small deviations obey the same laws as the deviations of β-rays, i e the number n out of a total number n_0 deviated through angles less than ϕ is given by $n/n_0 = 1 - \varepsilon^{-\phi^2/\overline{\phi^2}}$, where $\overline{\phi^2}$ is the average value of the squares of the deviations According to this formula the number deviated through angles much greater than the average deviation is quite negligible Thus if $\phi/\bar{\phi}$ is equal to 3 then $n/n_0 = 1 - 0 \cdot 00012$ It is found, however, that for large values of ϕ the number of α-rays scattered does not fall off with ϕ as rapidly as the theory of multiple scattering indicates, so that there is an appreciable number even with very large values of ϕ. To explain this Rutherford proposed the theory of *single scattering*, according to which the large deviations are due to single collisions with atoms and not to the summation of a large number of small deviations. Suppose α-rays to pass through a sheet consisting of a single layer of atoms, so that each ray passes through one atom, and suppose further that all but a very small fraction, say 1/1000, are only deviated through very small angles. Now let the rays pass through a second similar sheet. The chance that a particular α-ray will be deviated through more than a small angle by both sheets is only 1 in 1,000,000 and so is negligible. Even for a sheet consisting of 100 layers of atoms the chance that a ray is deviated once through more than a small angle is only about 1 in 10, and therefore the chance that a ray suffers two large deviations is about 1 in 100. In this way it is easy to see that the single-scattering theory should apply to the rays scattered through large angles by thin sheets, provided the total number scattered through large angles is only a small fraction of the whole number. According to this the scattering of α-rays through large angles by thin sheets will be the same as if the rays had all passed through one atom, and so will depend on the field of force inside the atom. To explain the large deviations an intense field in the atom is required, and this led Rutherford to propose his nucleus theory of the atom. According to this theory an atom consists of a positively charged nucleus surrounded by a number of electrons. In a neutral atom the positive charge on the nucleus is equal to the negative charge on the electrons, so that the charge on the nucleus is always an exact multiple of the charge e, which is a positive charge equal to the negative charge on one electron. Owing to the small mass of electrons it is clear that the α-rays cannot be deviated appreciably by them, so that the large deviations must be produced when an α-ray passes very close to a nucleus.

In fig. 18 let ABC be the path of an α-ray passing a nucleus at N Let AD be the path the ray would have followed in the absence of the nucleus, and let $\angle$ COD be the deviation φ. Let p be the length of the perpendicular on AD from N. We shall suppose the nucleus fixed in position at N. Let the co-ordinates of a point P on the path be NP $= r$ and $\angle$ PNO $= \theta$. The angular momentum of the α-ray about N is $mv_0p = h$, a constant, where v_0 is the initial velocity of the ray along AD, and m is the mass of the ray.

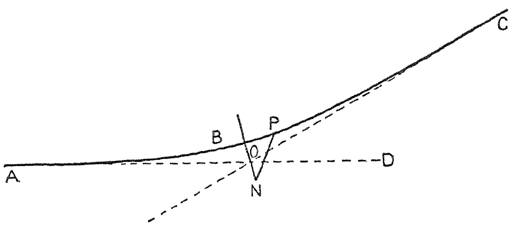


Fig 18

The kinetic energy of the ray is given by

$$\tfrac{1}{2}mv_0^2 - \tfrac{1}{2}mv^2 = 2Ee/r,$$

where v is the velocity of the ray when at a distance r from the nucleus, $2e$ is the charge of the α-ray, and E is the charge of the nucleus. The electrical potential due to the nucleus is E/r, so that the ray loses kinetic energy equal to $2Ee/r$ as it approaches the nucleus.

Now $v^2 = \dot{r}^2 + (r\dot{\theta})^2$, where $\dot{r} = dr/dt$ as usual. But $\dot{r} = \frac{dr}{d\theta}\dot{\theta}$; so that

$$v^2 = \left\{\left(\frac{dr}{d\theta}\right)^2 + r^2\right\}\dot{\theta}^2.$$

We have also $mv_0p = h = mr^2\dot{\theta}$, so that we get

$$v^2 = \left\{\left(\frac{dr}{d\theta}\right)^2 + r^2\right\}\frac{h^2}{m^2r^4}.$$

Now let $r = u^{-1}$, so that $\frac{dr}{d\theta} = -u^{-2}\frac{du}{d\theta}$, and

$$v^2 = \left\{u^{-4}\left(\frac{du}{d\theta}\right)^2 + u^{-2}\right\}\frac{h^2}{m^2}u^4.$$

Hence

$$\left(\frac{du}{d\theta}\right)^2 + u^2 = \frac{m^2}{h^2}\left\{-\frac{4Ee}{m}u + v_0^2\right\}.$$

The solution of this equation (cf. p. 393) is

$$u = A(\varepsilon\cos\theta - 1),$$

where A and ε are constants. Substituting this in the differential equation we find

$$A = \frac{Ee}{Tp^2} \text{ and } \varepsilon^2 - 1 = \left(\frac{Tp}{Ee}\right)^2,$$

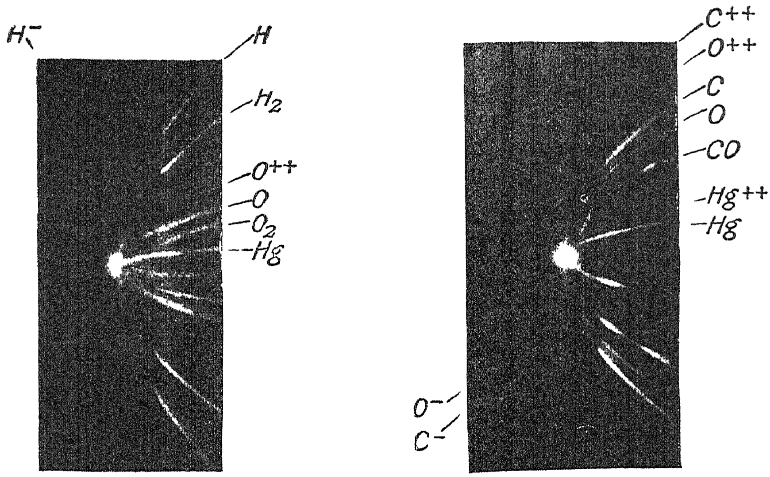
where $T = \frac{1}{2}mv_0^2$ is the initial kinetic energy of the α-ray.

In fig. 18, OA and OC are asymptotes to the path. Let AOB $=$ BOC $= \psi$, COD $= \varphi$

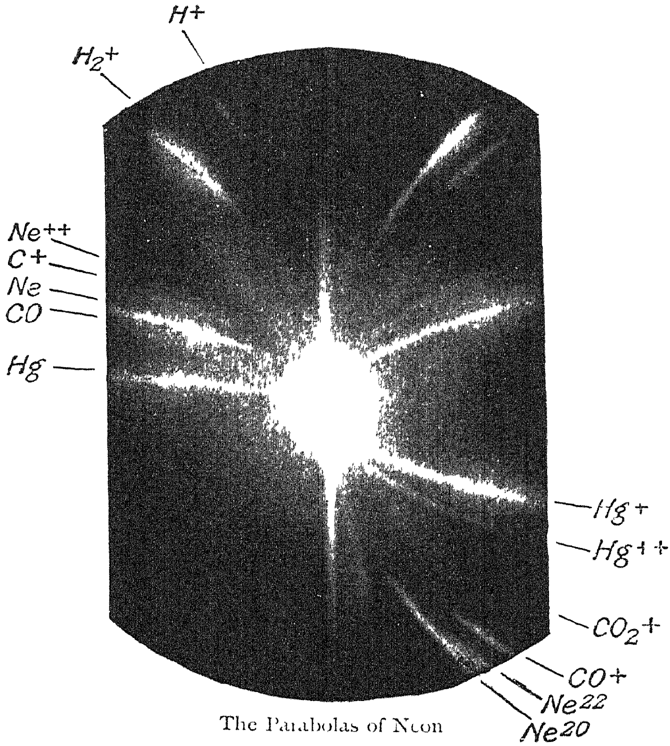
When r is very large we have $u = 0$, so that $\varepsilon\cos\psi = 1$, and we have then $\varphi + 2\psi = \pi$, or $\frac{1}{2}\varphi = \frac{1}{2}\pi - \psi$, and therefore $\sin\frac{1}{2}\varphi = 1/\varepsilon$, and $\cot\frac{1}{2}\varphi = \sqrt{\varepsilon^2 - 1}$, or

$$\cot\left(\frac{\varphi}{2}\right) = \frac{Tp}{Ee}.$$

Suppose now that a large number n of α-rays pass through a thin plate of thickness t and let the number of nuclei in the plate be N per unit volume The



Photographs of Typical Positive Ray Parabolas



The Parabolas of Neon

Fig. 2

After Aston *Isotopes* (Arnold)

Facing p. 223

total path of the rays in the plate is nt, so that the number of cases in which the undeviated path of a ray comes nearer to a nucleus than a distance p will be equal to the number of nuclei in a cylinder of length nt and cross-section πp^2. The number of such cases is therefore πntp^2N, and the number of rays which are deviated through an angle greater than φ is

$$\pi ntN\left(\frac{Ee}{T}\right)^2 \cot^2\left(\frac{\varphi}{2}\right).$$

The fraction of the rays deviated through angles greater than φ is therefore

$$\pi Nt\left(\frac{Ee}{T}\right)^2 \cot^2\left(\frac{\varphi}{2}\right).$$

We have not taken into account the charges on the electrons, which, of course, must limit the electric field due to the nucleus. However, the electrons are believed to be at distances from the nucleus of the order of 10^{-8} cm., and the values of p required to give large values of φ are much smaller than this, so that for the large deflections the field near the nucleus may be taken to be equal to E/r^2, as if the electrons were not present.

The number of α-rays scattered through different angles by thin plates of gold and other metals was found by passing a narrow nearly parallel beam of the rays through the plate and counting the scintillations produced on a small screen by the scattered rays. The small screen could be set up so as to receive the rays scattered through any desired angle. The total number of rays falling on the plate was also determined.

The results obtained agreed very closely with the single-scattering theory. The fraction of the rays scattered through an angle greater than φ was found to be proportional to the thickness of the scattering plate, inversely proportional to the square of the energy T of the rays, and proportional to $\cot^2(\varphi/2)$. From the values found for the fraction scattered it was possible to calculate the nuclear charge E for the atoms of the plate. It was found in this way, for example, that the nuclear charge for gold atoms is equal to $(79 \pm 0{\cdot}5)e$, in good agreement with the theory that the nuclear charge is equal to the product of the electronic charge by the atomic number, for the atomic number of gold is 79. Similar results were obtained with several other metals. These experiments therefore afford very strong support for Rutherford's nucleus theory of the atom.

22. α-ray Energies.

The energies of the α-rays from different radioactive bodies have been very exactly measured by finding the radius ρ of the circular path the rays describe in a uniform magnetic field. Such measurements have been made by Rosenblum, Rutherford and his associates, and by Briggs. The apparatus used by Rutherford is shown in fig. 19. The magnetic field was produced by a large electromagnet having annular poles indicated by the two concentric circles. The magnetic field was perpendicular to the plane of the diagram and nearly uniform between the two circles. The source of the α-rays was at S. Some of the rays from the source describe circular paths through a diaphragm at D and are focussed on a detector at P. The diameter of the paths SP was 80 cm. The magnetic field was adjusted so as to focus the rays on the detector. The source was usually a metal

plate 8×3 mm. coated with the radioactive deposit on one side and placed at S with the active side nearly parallel to the circular ray paths. The detector was a tube with a narrow slit covered with thin mica through which the rays passed. The tube contained two plane electrodes and gas at a few centimetres pressure. An α-ray ionizes the gas, and the ions are collected on the electrodes and slightly change the potential difference between them. The potential change is amplified and recorded by an impulse counter. In this way the number of

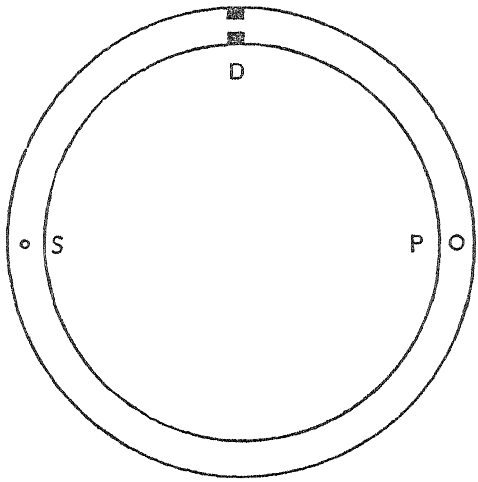


Fig 19

α-rays entering the slit of the detector can be counted. The magnetic field is varied by small steps and the number of rays counted. The number counted rises to a sharp maximum when the field has certain values.

If N denotes the number of atoms in a gramme atom, that is a number of grammes equal to the atomic weight W, and F the faraday, then $Ne = F$ and $NM_0 = W$, so $e/M_0 = F/W$, where e is the protonic charge and M_0 the rest mass of the atom in question. For helium atoms with charge $2e$ we have $W = W_{He} - 2w_0$, where W_{He} is the atomic weight of a neutral helium atom and w_0 that of an electron, so $2e/M_0$ for α-rays is $2F/(W_{He} - 2w_0)$. If the rays describe a circle of radius ρ in a magnetic field H, then $2Hev = Mv^2/\rho$, where $M = \dfrac{M_0}{\sqrt{1 - v^2/c^2}}$, so

$$\frac{v}{H\rho} = \frac{2e}{M_0}\sqrt{1 - v^2/c^2}.$$

The energy P of the α-rays, in electron volts, is then given by

$$P = \frac{c}{10^8}\left\{\left(\left(\frac{M_0 c}{e}\right)^2 + (2H\rho)^2\right)^{1/2} - \frac{M_0 c}{e}\right\}.$$

The energy set free by the disintegration when the α-ray is emitted is equal to the kinetic energy of the α-ray plus that of the recoil of the radioactive atom. If M_R is the mass of the atom which recoils and v_R its velocity of recoil, then $M_R v_R = Mv$, so that v_R can be easily calculated and the energy $\frac{1}{2}M_R v_R{}^2$ obtained.

The following table gives the energies, in electron volts, of several α-rays, also their ranges in centimetres in air at 15° C. and 760 mm. pressure, and the disintegration energies.

Element.	Mean Range.	Energy $\times 10^{-6}$.	Disintegration Energy $\times 10^{-6}$.
Radon	4 014	5 488	5·589
RaC′	6 870	7 683	7·829
,,	9 00	9 069	9·241
Polonium	3 805	5·300	5·403
Radio-Ac	—	6·051	6·159
,,	—	5 674	5·776
Thorium C′	8 533	8·778	8 947
,,	9 687	9 491	9·674
Thorium-C	—	6 044	6·160

The α-ray energies, like the γ- and secondary β-ray energies, form many pairs with sums equal to multiples of $q = 3{\cdot}85 \times 10^5$ electron volts. The energies are therefore equal to $nq \pm C_k$, and the C_k's have values equal to the γ- and β-ray values. This indicates that the radioactive atoms probably have equally spaced energy levels with the constant difference $3{\cdot}85 \times 10^5$ electron volts.

23. Theory of α-ray Disintegrations.

As we have seen, the single scattering of α-rays by heavy elements, like gold, agrees with the theory that the potential energy of the α-ray, due to the repulsion of the nuclear charge Ze, is equal to $2Ze^2/r$, where r is the distance between the centres of the particles. In the case of uranium, α-rays with energy 9 MEV are scattered in accordance with the theory. This shows that the electric field of the uranium nucleus is equal to Ze/r^2 for values of r equal to and greater than the value for which $2Ze^2/r = 9\ MEV$. But the uranium nucleus emits α-rays with energy only about 4 MEV, so that on the classical theory these rays must have started from a point outside the nucleus where $2Ze^2/r = 4\ MEV$, which is clearly impossible. This difficulty is removed by the quantum mechanics theory, according to which a particle inside the nucleus may escape with energy much less

than its potential energy when just outside the nucleus. This application of quantum mechanics was first worked out by Gurney and Condon and by Gamow, independently, at the same time. It is supposed that the potential energy P of an α-ray is a function of its distance r from the centre of the nucleus of the form OGCFD shown in fig. 20. For values of r greater than $OG = a$ the potential energy is equal to $2Ze^2/r$, represented by the curve CFD, and for values of r less than OG the potential energy is constant. The radius a may be regarded as that of the nucleus. An α-particle with total energy E, represented by the height of the dotted line HEFK, would have negative kinetic energy for values of r between HE and HF, and so could not pass through this region on the classical theory.

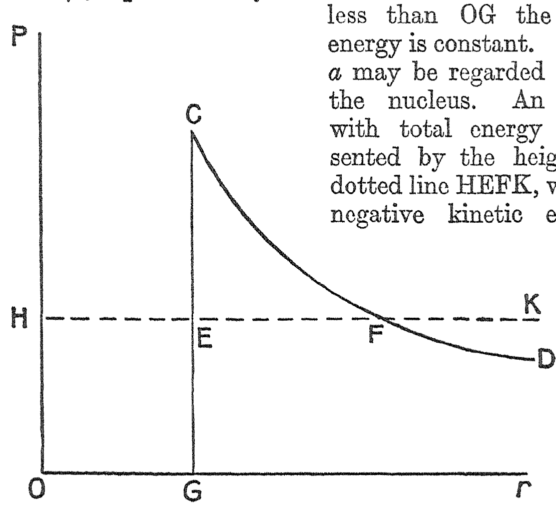


Fig 20

Schrödinger's equation for the particle between $r = OG = a$ and $r = HF = b$ is

$$\Delta w + \frac{8\pi^2 m}{h^2}\left(E - \frac{2Ze^2}{r}\right)w = 0.$$

Let $\psi = wr$ and suppose ψ is a function of r only, so that

$$\frac{d^2\psi}{dr^2} + \frac{8\pi^2 m}{h^2}\left(E - \frac{2Ze^2}{r}\right)\psi = 0.$$

Now let $\psi = e^S$, where S is a function of r only, so that

$$\frac{d^2\psi}{dr^2} = e^S\left(\frac{d^2S}{dr^2} + \left(\frac{dS}{dr}\right)^2\right)$$

and

$$\frac{d^2S}{dr^2} + \left(\frac{dS}{dr}\right)^2 + \frac{8\pi^2 m}{h^2}\left(E - \frac{2Ze^2}{r}\right) = 0.$$

For an approximate solution we may neglect $\frac{d^2S}{dr^2}$, so that between $r=a$ and $r=b$, where $E-2Ze^2/r$ is negative we have $\frac{dS}{dr}=\pm 2\pi p/h$, where $p=\sqrt{2m\left(\frac{2Ze^2}{r}-E\right)}$, which gives

$$S_a - S_b = \pm\frac{2\pi}{h}\int_a^b p\,dr.$$

The positive sign must be taken when we are supposing that an α-particle is in the nucleus, since S_b will determine the chance of its being found at $r=b$, which must be much less than the chance of finding it at $r=a$. If we suppose the α-particle inside the nucleus is moving backwards and forwards along a radius with velocity v, then it will arrive at $r=a$, $v/2a$ times per second. If it escapes it will arrive at $r=b$, and since the chance of finding it at $r=b$ is to the chance of finding it at $r=a$ as $\bar{\psi}_b\psi_b$ is to $\bar{\psi}_a\psi_a$, the number of times it escapes in a second will be

$$\frac{v}{2a}\frac{\bar{\psi}_b\psi_b}{\bar{\psi}_a\psi_a}=\frac{v}{2a}\frac{e^{2S_b}}{e^{2S_a}}=\frac{v}{2a}e^{-\frac{4\pi}{h}\int_a^b p\,dr}$$

The average life τ of the α-particle inside the nucleus will be the reciprocal of this, so that

$$\tau=\frac{2a}{v}e^{\frac{4\pi}{h}\int_a^b p\,dr},$$

or if $\frac{1}{2}mv^2=E$ so that $v=\sqrt{2E/m}$, then

$$\tau=a\sqrt{\frac{2m}{E}}\,e^{\frac{4\pi}{h}\int_a^b p\,dr}.$$

In the integral $p=\sqrt{2m\left(\frac{2Ze^2}{r}-E\right)}$ and $b=2Ze^2/E$, so that

$$\int_a^b p\,dr=\sqrt{2mE}\int_a^b\sqrt{b/r-1}\,.\,dr.$$

Substituting $r=b\sin^2\theta$, we get $2b\sqrt{2mE}\int\cos^2\theta\,.\,d\theta$, which gives

$$\int_a^b p\,dr=b\sqrt{2mE}\left\{\cos^{-1}\sqrt{\frac{a}{b}}-\sqrt{\frac{a}{b}\left(1-\frac{a}{b}\right)}\right\},$$

so that

$$\tau=a\sqrt{\frac{2m}{E}}\,e^{\frac{4\pi}{h}b\sqrt{2mE}\left\{\cos^{-1}\sqrt{\frac{a}{b}}-\sqrt{\frac{a}{b}\left(1-\frac{a}{b}\right)}\right\}}.$$

This result shows that τ is a function of the energy E of the α-ray, the atomic number Z, and the nuclear radius a.

If the values of E and τ for the naturally radioactive bodies, which emit α-rays, are put in this equation for τ, the values of the nuclear radius a can be calculated. In this way values of a about 10^{-12} cm. are obtained, so that since a is probably about the same for all the naturally radioactive bodies it appears that the above theory, roughly speaking, gives correct results.

The experimental values of E and τ show that $\log \tau = A - B \log E$ approximately, where A and B are constants. This result, known as Geiger and Nuttall's law, is, roughly speaking, in agreement with the above theoretical value of τ, assuming the radius a to have the same value for all the naturally radioactive bodies.

REFERENCES

1. *Conduction of Electricity through Gases.* J. J. Thomson.
2. *Electricity in Gases.* J. S Townsend.
3. *Radioactivity.* E. Rutherford.

CHAPTER X

Positive Rays

1. Nature of Positive Rays.

Positive rays are positively charged atoms or molecules which have acquired a high velocity in an electric field. They are produced in the electric discharge in gases at low pressures and in other ways. If a discharge is passed between two electrodes in a gas at a low pressure the gas is ionized, that is, electrons are removed from the gas molecules, which therefore become positively charged. These positive ions move towards the negative electrode, and when the potential difference is large, say 50,000 volts, they acquire a high velocity. If a small hole is bored through the negative electrode, some of the positive ions pass through the hole and form a beam of positive rays. They can be deflected by transverse magnetic and electric fields. The deflections are much smaller than those of cathode rays or electrons and are in the opposite direction, showing that the rays are positively charged.

These rays have been investigated by Wien, J. J. Thomson, Aston, and others, and very important results have been obtained. The positive rays cause gases through which they pass to emit light, and they produce phosphorescence when they fall on solid bodies. The mineral willemite, a silicate of zinc, phosphoresces brightly when exposed to positive rays and was found by J. J. Thomson to be one of the best substances for locating a beam of the rays. They affect a photographic plate like light, and so measurements of their deflections are usually made by photographing them.

2. Experiments of J. J. Thomson.

A discharge tube used by J. J. Thomson is shown in fig. 1. It consists of a large glass bulb BB of about 1500 c. c. capacity, having a small electrode sealed in at D. The cathode C consists of an iron tube about 7 cm. long with an aluminium cap on it at C. Along the axis of this tube a very fine straight copper tube is mounted, through which the rays pass into a conical tube K where they fall on a willemite screen or photographic plate P. The narrow beam of rays is passed between two insulated iron blocks E and F which can be magnetized by the poles G, H of a large electromagnet, so producing a magnetic

field between E and F perpendicular to the beam of rays. The blocks E and F can be connected to a battery so as to produce an electric field between them in the same direction as the magnetic field. The bulb B and tube K are connected to pumps by means of which they can be exhausted

A nearly perfect vacuum is maintained in K, and the gas to be examined is slowly admitted to B through a narrow tube, the pressure in B being kept constant at a suitable small value by the pumps. The electrode D is charged positively by a large induction coil and the cathode C connected to the earth. When the pressure in B is low the potential difference required to produce a discharge in the bulb is large and positive rays are produced, some of which pass through the narrow tube and give a small spot on the plate P. When the magnetic and electric fields between E and F are produced the rays are deflected and distributed on the plate P along parabolic curves, the positions of which depend on the charges, masses, and velocities of the rays.

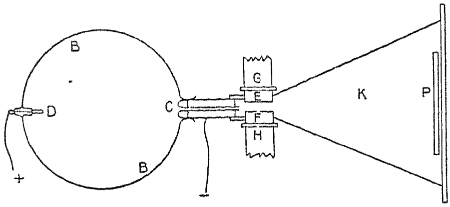


Fig 1

3. Theory of the Positive Ray Parabolas.

The force on a particle of mass m, carrying an electric charge e, moving with velocity v in a magnetic field H perpendicular to v, is Hev and is perpendicular to H and v. If then the particle travels a distance d in the magnetic field it will acquire a velocity u perpendicular to the field given by $mu = Hevd/v$, so that $u = Hed/m$. The deflection of the particle after travelling a distance l beyond the field will be approximately $y = lu/v = Hedl/mv$.

The force on the particle in an electric field F is Fe, so that while travelling a distance d in an electric field F perpendicular to v it will acquire a velocity w in the direction of F given by $mw = Fed/v$. The deflection z after going a further distance l, where $F = 0$, is given by $z = lw/v = Fedl/mv^2$.

In J. J. Thomson's apparatus H and F were both perpendicular to the faces of the blocks E and F, and both extended over the same distance d, so that the electric and magnetic deflections were at right angles to each other and equal to y and z The positive rays having a given value of e/m but different velocities v therefore fall on a parabola on the plate P, the equation of which is got by eliminating v from

$$y = \frac{Hedl}{mv} \text{ and } z = \frac{Fedl}{mv^2}.$$

The equation of the parabola is therefore

$$\frac{Fedl}{mz} = \left(\frac{Hedl}{ym}\right)^2,$$

or

$$y^2 = \frac{H^2}{F}\frac{edl}{m}z.$$

By measuring y and z for points on the parabola, e/m can be determined when H, F, d, and l are known

When a mixture of gases is put in the discharge tube and the positive rays from it examined in this way a number of distinct parabolas is obtained on the plate P, one for each value of e/m of the positive rays produced.

The ratio of y to z is given by

$$\frac{y}{z} = \frac{H}{F}v,$$

so that all rays having the same velocity but different values of e/m fall on a straight line passing through the origin ($y = 0$, $z = 0$) and inclined to the z axis at an angle θ given by

$$\tan\theta = \frac{H}{F}v.$$

Fig. 2 shows several sets of parabolas obtained on photographic plates by J. J. Thomson. The letters at the ends of the parabolas indicate the atom or molecule to which the parabola is attributed. If the charge on the atom is equal to that on an atom which has lost one electron or e, this is either not indicated or indicated by one plus sign, and if the charge is $2e$ this is indicated by two plus signs, thus Hg^{++} indicates the parabola attributed to mercury atoms each having a charge $2e$.

If m_H is the mass of one hydrogen atom, m_R the mass of one positive ray, and e_H and ne_H are the charges they carry, then $e_H/m_H = 9650$ electromagnetic units per gramme, so that if ne_H/m_R, which is e/m for the positive rays, is equal to x, then

$$\frac{9650}{x} = \frac{m_R}{nm_H}.$$

If the positive ray consists of N atoms of atomic weight A, then $NA = m_R/m_H$ very nearly, so that $x = 9650n/NA$.

For example, in the case of mercury, $A = 200$, and we get the following possible values of e/m or x for single mercury atoms

n		$e/m = x$
1	...	48·25
2	...	96·5
3	...	144·75
4	...	193·0
5	...	241·25
6	...	289·50
7	....	337·75

It was found that with mercury vapour seven distinct parabolas could be obtained corresponding to the values of e/m given above. Thus it appears that mercury atoms can lose from one to seven electrons in the discharge tube.

In hydrogen, parabolas corresponding to $n = 1$, $N = 1$; $n = 1$, $N = 2$; and $n = 1$, $N = 3$ were obtained.

The parabolas of many elements and compounds have been observed and the corresponding values of e/m are always such as can be readily explained by supposing the rays to consist of an atom or molecule of the substance in question which has lost one or more electrons.

4. Have All the Atoms of a Given Element the Same Mass?

This method of positive ray analysis gives a definite answer to the question whether all the atoms of a given element have equal

masses or not. If well defined parabolas are obtained they must be due to particles all having the same value of e/m and so presumably of m. If the masses of the atoms of an element varied between certain limits above and below a mean value equal to that corresponding to the chemical atomic weight, then well defined parabolas would not be obtained but more or less broad diffuse ones. It is clear, therefore, that this is not the case However, the possibility remains that the atoms of a given element are not all of the same mass, but that there are two or more possible masses each of which gives a well defined parabola. It is found that such is the case. J. J. Thomson obtained two parabolas with neon corresponding to atomic weights 20 and 22, whereas the accepted atomic weight of neon is 20·2. Aston and G. P. Thomson found that lithium also gives two parabolas corresponding to atomic weights 6 and 7, whereas the chemical atomic weight is 6·94. G P. Thomson has examined several light elements by the parabola method. He found beryllium to be a simple element.

5. Aston's Method of Positive Ray Analysis.

A new method of positive ray analysis was invented by Aston which enables e/m to be determined with much greater accuracy than by the parabola method. Aston's apparatus is shown in fig 3.

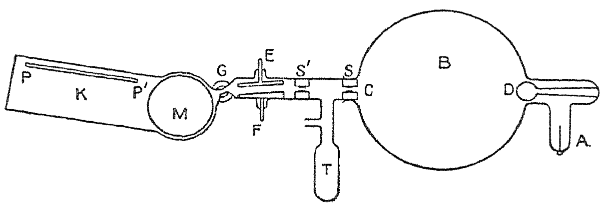


Fig 3

The discharge producing the positive rays takes place in a large bulb B between a concave cathode C and an anode A. The cathode rays from C fall on a quartz bulb D and are thus prevented from fusing the glass wall of the discharge tube. In the cathode C there is a narrow slit S through which the positive rays pass. A second slit at S′ allows a narrow beam of the rays from S to pass. The space between S and S′ is kept highly exhausted by means of a bulb T containing charcoal cooled in liquid air. The beam of rays passes between two parallel metal plates E and F between which an electric field is maintained. The rays are deflected by this field and the narrow beam is spread out into a broad band. Part of this band is allowed to pass through an adjustable slit G. The rays then pass between the poles of a magnet M and are deflected by the magnetic field in the opposite direction to the previous electric deflection. Finally, they pass into

a box K and fall on a photographic plate PP′. This arrangement causes all rays having the same value of e/m to strike the plate at the same point, so that lines are obtained on the plate one for each value of e/m. The series of lines on the plate is called a "mass spectrum" and the apparatus a "mass spectrograph". The box K and the rest of the apparatus except the bulb B must be highly exhausted, since any appreciable amount of gas stops the positive rays. The gas to be examined is slowly admitted to B and the pressure kept constant by pumps as in J. J. Thomson's apparatus.

6. Theory of Aston's Mass Spectrograph.

Let the angle through which the rays are deflected by the electric field be θ and by the magnetic field ϕ. Also let b be the distance from the middle of the electric field to the middle of the magnetic field.

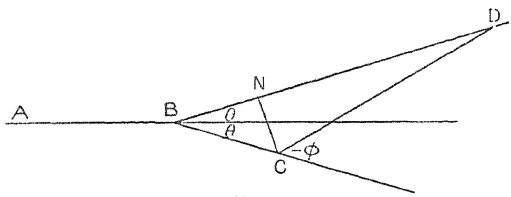


Fig 4

Consider two rays for which e/m is the same but which have velocities v and $v + dv$ respectively. Then we have approximately

$$\theta = \frac{Fed}{mv^2}, \text{ and } \phi = \frac{Hed'}{mv},$$

as in J. J. Thomson's apparatus, where d is the length of the electric and d' that of the magnetic field. Hence

$$d\theta = -2\frac{Fed}{m}\frac{dv}{v^3}, \quad d\phi = -\frac{Hed'}{mv^2}dv,$$

so that

$$\frac{d\theta}{\theta} = -2\frac{dv}{v}, \text{ and } \frac{d\phi}{\phi} = -\frac{dv}{v},$$

and therefore

$$d\phi = \frac{\phi}{2\theta}d\theta.$$

The separation of the two rays at the magnetic field is $bd\theta$, and at a distance r farther on it is $bd\theta + r(d\theta + d\phi)$ or $d\theta\left\{b + r\left(1 + \frac{\phi}{2\theta}\right)\right\}$.

This is zero when

$$b + r\left(1 + \frac{\phi}{2\theta}\right) = 0,$$

or when

$$r = \frac{-b}{1 + \phi/2\theta}.$$

This requires ϕ to be negative and greater than 2θ for r to be positive.

In fig. 4 let AB be the original direction of the positive rays and let them be electrically deflected at B through the angle θ along BC.

Let the magnetic deflection take place at C through the angle $-\phi$ greater than 2θ so that they go along CD. Draw BND making an angle θ with AB produced and let CD and BND meet at D. Then the angle at D is equal to $-(\phi + 2\theta)$, so that

$$NC = b \sin 2\theta = -r \sin(\phi + 2\theta),$$

where $b = BC$, and $r = CD$. If θ and ϕ are small angles then approximately

$$2b\theta = -r(\phi + 2\theta),$$

or

$$r = \frac{-b}{1 + \phi/2\theta}.$$

Thus we see that rays having the same value of e/m but slightly different velocities will all cross BND at the same point. In Aston's apparatus the photographic plate is put along BD, so that all the rays having a definite value of e/m fall on the plate at the same point. If rays are present having several values of e/m a series of lines is produced on the plate, one for each value of e/m. By varying the electric and magnetic field strengths rays having any desired range of values of e/m can be brought on to the plate. In this way Aston got well defined lines on his plates, and was able to compare values of e/m to one part in one thousand by measuring the distances between the lines.

7. Aston's Results.

Aston has examined the positive rays from many different elements and finds that the relative masses of the atoms can be expressed by integers within the limits of error. For example, the masses of the rays from carbon, nitrogen, oxygen, and fluorine are proportional to 12, 14, 16, and 19 to within one part in 1000. These numbers agree with the atomic weights 12·00, 14·008, 16, and 19·00. Hydrogen rays form an exception to this rule, for, taking $O = 16$, hydrogen rays give $H = 1{\cdot}008$, in agreement with the chemical atomic weight.

The rays of elements which have atomic weights differing appreciably from integers are found to contain atoms having two or more different masses proportional to integers. For example, chlorine gives rays with masses proportional to 35 and 37, taking oxygen rays proportional to 16 as the standard. The atomic weight of chlorine is 35·46, so that it appears that chlorine is a mixture of two sorts of chlorine atoms having atomic weights 35 and 37. Xenon, the atomic weight of which is 130·2, is found to consist of a mixture of atoms of atomic weights 128, 129, 130, 131, 132, 134, and 136. In this way most of the chemical elements have been found to consist of mixtures of atoms of different atomic weights.

8. Hot Anode Method for Positive Rays.

The discharge tube method of obtaining positive rays is only successful in the case of elements which form stable gaseous compounds which can be admitted to the tube. Positive rays from many very slightly volatile substances can be obtained by means of a hot anode on which a small quantity of the substance is placed. The hot anode is usually a narrow strip of platinum foil which can be heated by passing a current through it. The substance or mixture to be investigated is put on the strip of foil and fused on to it by heating the platinum. The hot anode is mounted directly in front of the cathode of the mass spectrograph and a very good vacuum is maintained in the bulb. When a large potential difference is maintained between the anode and the cathode and the anode is heated, positive rays are emitted by the anode and some of them pass through the slit in the cathode.

The ratio e/m for alkali metal atoms emitted by alkali sulphates on such a hot anode was determined by O. W. Richardson by measuring their deflections in an electric and transverse magnetic field. Richardson found values of e/m agreeing closely with those calculated for single atoms which have lost one electron. His experiments were not sufficiently accurate to distinguish between atomic weights differing by only a few units.

Aston and G. P. Thomson examined the positive rays of lithium by the parabola method and got two parabolas corresponding to atomic weights 6 and 7. Aston examined the other alkali metals, using a hot anode and his mass spectrograph. He found sodium to give rays of atomic weight 23 only. Potassium gave rays with atomic weights 39 and 41. Rubidium gave rays of weights 85 and 87, and cæsium gave only rays of atomic weight 133.

9. Dempster's Method.

The positive rays from several elements have been studied by Dempster by another method. In his apparatus positive rays from a hot anode are accelerated by a known potential difference and then pass through a slit. Immediately beyond the slit they enter a uniform magnetic field which causes them to move along a semicircular path into another slit. The rays passing through the second slit are received on an insulated electrode and the charge they carry is measured.

If P is the potential difference used, then

$$Pe = \tfrac{1}{2}mv^2,$$

where e is the charge, m the mass, and v the velocity of the rays due to

P. If H is the strength of the magnetic field, and r the radius of the semicircular path, then

$$\frac{mv^2}{r} = Hev.$$

Hence

$$\frac{m}{e} = \frac{H^2 r^2}{2P}.$$

Dempster kept H constant, and measured the current received by the insulated electrode as P was increased by small steps. It was found that the current increased to a maximum at certain values of P, and the corresponding values of m/e were calculated. In this way Dempster found magnesium to consist of a mixture of three sorts of atoms having atomic weights 24, 25, and 26. He also found zinc to consist of atoms with weights 63, 65, 67, and 69.

10. Isotopes.

It was formerly supposed that all the atoms of an element are precisely equal in all respects, but the methods used to determine atomic weights depended on the relative weights of quantities which contained enormous numbers of atoms. The chemical atomic weights are therefore merely the average atomic weights for very large numbers of atoms. The methods of positive ray analysis first made it possible to test the question whether all the atoms of an element are equal or not.

It is not possible to separate the two sorts of chlorine atoms with atomic weights 35 and 37 by any chemical process. They have precisely similar chemical properties and are therefore regarded as atoms of the same element. The same is true of the different sorts of atoms of other elements. Substances having the same chemical properties but different atomic weights are called *isotopes*.

In a few cases it has been found possible by physical processes such as diffusion or evaporation to slightly alter the relative proportions of the isotopes in certain elements. In this way a quantity of the element is separated into parts having slightly different atomic weights. This was done for neon by Aston, for chlorine by Harkins, and for mercury by Bronsted and Hevesy. The last named physicists prepared two samples of mercury having relative densities 0·99974 and 1·00023 compared with ordinary mercury taken as unity.

The theory of isotopes is discussed in the chapter on Atomic Nuclei. The important application of positive ray analysis to ionization by collisions of electrons, by Smyth, is discussed in the chapter on the Critical Potentials of Atoms.

11. Precision Mass Spectrographs.

In 1925 Aston designed an improved mass spectrograph believed to be capable of measuring the relative masses of positive rays to about

1 in 10,000. The new instrument was similar to the old one but larger, and the angles through which the rays were deflected were about doubled.

In 1936 Dempster and also Bainbridge, working independently, constructed new mass spectrographs of improved and rather similar designs. The new instruments replaced older less exact ones with which they had previously obtained valuable results. Bainbridge's new mass spectrograph is shown in fig. 5.

Positive ions are produced in the tube A by passing a current at about 15,000 volts through neon at a low pressure. The element to

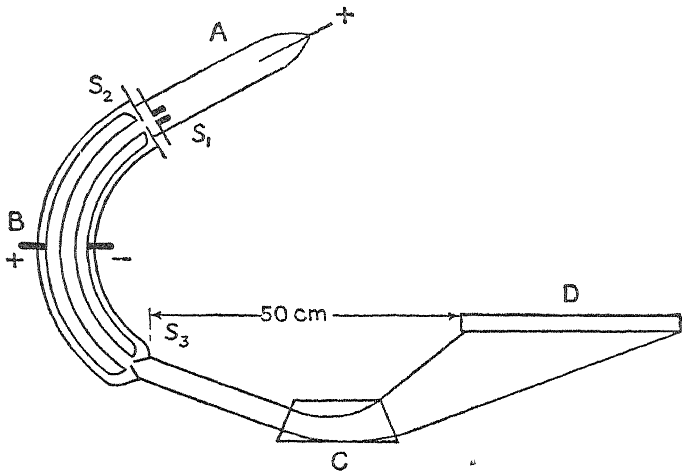


Fig 5

be investigated is put on the cathode S_1 and is vaporized by the positive neon ions. Positive ions from this discharge in A pass through two narrow slits S_1 and S_2, and are deflected by a radial electric field between the plates of a cylindrical condenser B. The condenser deviates the rays through an angle approximately $\pi/\sqrt{2}$ on to a third slit at S_3. Rays diverging from S_2 and all having the same kinetic energy are focussed on the plane through S_3, so that an energy spectrum is formed at S_3. This method of electrostatic focussing is due to Hughes and Rojansky. The rays which pass through S_3 are deviated by a magnetic field at C and fall on a photographic plate at D. Rays with equal m/e values are focussed on the plate at D as in Aston's mass spectrographs. The variation of m/e along the plate is very nearly linear, and with narrow slits a resolving power of 10,000 or more is obtained. This means that the lines on the plate due to ions with m/e values differing by one in 10,000 are separated appreciably. The

lines due to $_1H^2$ and 2_1H^1 which differ by 0·0014 atomic weight units are separated by about ten times the width of each line.

The mass spectrograph enables very exact values of the differences between two nearly equal atomic masses to be found. From such small differences the atomic masses with $O^{16} = 16$ can be determined. This is called the doublet method. For example, the following mass differences have been found;

$$H_2 - D = 0{\cdot}00153,$$
$$D_3 - \tfrac{1}{2}C^{12} = 0{\cdot}04219,$$
$$C^{12}H_4 - O^{16} = 0{\cdot}03649.$$

These three equations give $H = 1{\cdot}00813$, $D = 2{\cdot}01473$ and $C^{12} = 12{\cdot}00398$ with $O^{16} = 16$.

The following table gives the atomic weights, with $_8O^{16} = 16$, of a number of elements found by means of precision mass spectrographs.

Element.	Atomic Weight.
H^1	1·00813
H^2	2 01473
He^4	4 00389
Li^7	7·01818
Be^9	9·01516
B^{10}	10·01631
B^{11}	11·01292
C^{12}	12 00398
N^{14}	14·00750
Ne^{20}	19·99881
Ne^{22}	21·99864

REFERENCES

1. *Rays of Positive Electricity*. J. J. Thomson.
2. *Isotopes*. F. W. Aston.

CHAPTER XI

Radioactive Transformations

1. Discovery of Radioactivity. Radium.

Radioactivity was discovered by Becquerel in 1896. Röntgen's discovery in 1895 of rays which can penetrate materials opaque to ordinary light, and which were apparently associated in some way with the fluorescence of the glass of the Crookes tubes from which they were emitted, suggested the possibility that phosphorescent or fluorescent substances might emit such rays. Becquerel placed a phosphorescent compound of uranium—the double sulphate of uranium and potassium—on a photographic plate wrapped in opaque black paper. He found that the plate was affected and that rays were emitted by the uranium compound which penetrated the paper and other substances. By putting metallic objects between the uranium and the plate shadows of the objects were obtained on the plate.

It was soon found that all compounds of uranium, whether phosphorescent or not, emit the rays, and that the activity depends on the amount of uranium present in the compound and not at all on the other elements present. Many other substances were then tested in the same way. Mme Curie in 1898 found that thorium is also radioactive, and that the mineral pitch-blende and other uranium minerals are four or five times more active than would be expected from the amount of uranium in them. This suggested that uranium minerals contain another radioactive body more active than uranium, and Curie found as the result of a laborious chemical separation of large quantities of pitch-blende residues that there is present in these a substance which has chemical properties similar to barium and which is enormously more radioactive than uranium. This substance was called radium. Another radioactive body, similar to bismuth, was also discovered, and was called polonium.

Curie found that there is about 0·35 gm. of radium to 1 ton of uranium in uranium minerals Radium is more than one million times more radioactive than uranium. It is not easily separated from barium. Radium bromide is less soluble in water than barium

bromide, which enables them to be separated by a series of crystallizations

Radium is found to have all the properties of a metallic element. Its atomic weight is 226·0 and it occupies a place in the periodic table of elements in the same column as calcium, strontium, and barium. It gives a spectrum similar to that of the other alkaline earth metals.

Rutherford in 1899 found that the rays from radioactive bodies are of two kinds, which he called α- and β-rays. The α-rays are much less penetrating than the β-rays. Villard in 1900 discovered that a third type of rays more penetrating than the β-rays is also emitted, which were called γ-rays. The properties of these three kinds of rays are discussed in other chapters, and it will suffice here to say that α-rays are positively charged helium atoms shot out with velocities about $\frac{1}{10}$ that of light, β-rays are high-velocity electrons, and γ-rays are electromagnetic waves of extremely short wave-length,

2. **Uranium-X and Thorium-X.**

An important discovery was made by Sir William Crookes in 1900. He found that if ammonium carbonate is added to a solution of uranium nitrate until the precipitate first formed is redissolved, a small quantity of insoluble precipitate remains which can be filtered off. This small precipitate Crookes found to be strongly radioactive when tested by its action on a photographic plate, whereas the uranium solution is no longer active. The radioactivity of the uranium is therefore due to the presence of a small quantity of another substance, intensely active, which was called uranium-X. It was found later that the uranium-X emits only β-rays, and that the uranium emits α-rays. α-rays have not much action on a photographic plate as compared with β-rays, which explains Crookes' results.

The activity of the uranium-X was soon found not to be permanent but to die away gradually so that after a few months it disappeared, whereas the β-ray activity of the uranium gradually recovered after the separation of the uranium-X and after a few months was completely restored. A second quantity of the uranium-X could then be separated from the uranium. It appears therefore that uranium continually produces uranium-X, the activity of which gradually dies away.

Rutherford in 1900 discovered that thorium emits a radioactive gas which was called thorium emanation. The radioactivity of this gas was found to die away in a few minutes. In 1902 Rutherford and Soddy separated an intensely radioactive body from thorium similar to uranium-X, which was called thorium-X. It was found that thorium-X emits the thorium emanation.

M. and Mme Curie discovered that radium also emits an emanation,

and that solid bodies exposed to this emanation become radioactive. The emanation deposits an active material on bodies in contact with it. Rutherford shortly afterwards found that thorium emanation also gives a radioactive deposit. The activity of these active deposits eventually disappears.

3. Rutherford and Soddy's Theory.

A theory of radioactivity, now universally accepted, was put forward by Rutherford and Soddy in 1902 to explain the results they had obtained with thorium and its products. According to this theory the atoms of radioactive elements are unstable and eventually explode or decompose into new atoms having quite different properties. The chance that a particular atom will decompose has a definite value, so that if we consider a large number of atoms the number which decompose per unit time is proportional to the number existing. Thus if there are N atoms at time t, then

$$\frac{dN}{dt} = -\alpha N,$$

where α is a constant, exactly as for a monomolecular chemical decomposition. Integrating this equation we get

$$N = N_0 \varepsilon^{-\alpha t},$$

where N_0 is the number of atoms existing at the time $t = 0$. It was found that thorium-X and thorium emanation disappear gradually in accordance with this equation. When $(N/N_0) = \frac{1}{2}$ we have $t = (\log_e 2)/\alpha$, where t is the time in which N falls to $N/2$. The rate of disappearance of radioactive bodies is conveniently expressed by giving the value of $(\log_e 2)/\alpha$, i.e. the time in which the amount remaining drops to one-half the initial value. For thorium emanation this time is 54 sec., and for thorium-X it is 3·64 days. For uranium-X the half value period is 24·6 days.

When a radioactive atom decomposes, the products formed are an α-ray or a β-ray, but not both, and a new atom. When an α-ray is one of the products the new atom has an atomic weight four units less than that of the parent atom, because the α-ray is a positively charged helium atom of atomic weight 4. When a β-ray is emitted the atomic weight of the new atom is almost exactly equal to that of the parent atom, since the mass of a β-ray or electron is less than one thousandth part of that of one hydrogen atom.

According to the nucleus theory of atoms an atom consists of a positively charged nucleus surrounded by electrons. It is supposed that it is the nucleus which decomposes in radioactive changes. Thus when an α-ray is emitted the charge on the nucleus is diminished by $2e$, the charge on the α-ray, and when a β-ray is emitted the nuclear charge is increased by e, since $-e$ is the charge on one electron. Thus when an α-ray is emitted the atom must also lose two electrons if it remains neutral, and when a β-ray is emitted it must gain one more electron to remain neutral.

Let us suppose that at time $t = 0$ we have N atoms of a radioactive substance and that these atoms decompose, giving a second radioactive substance, which decomposes giving a third substance, and so on until finally a non-radioactive product is produced. Let N_1, N_2, N_3, . . . N_n, be the numbers of the atoms of the successive products existing at time t, N_n being the number of atoms of the final product.

Then we have

$$\frac{dN_1}{dt} = -\alpha_1 N_1,$$

$$\frac{dN_2}{dt} = \alpha_1 N_1 - \alpha_2 N_2,$$

$$\frac{dN_3}{dt} = \alpha_2 N_2 - \alpha_3 N_3,$$

$$\cdots\cdots\cdots$$

$$\frac{dN_n}{dt} = \alpha_{n-1} N_{n-1}.$$

The sum of all the terms on the left-hand sides is zero, showing that the total number of atoms remains constant, which is obviously correct. The first equation gives $N_1 = N\varepsilon^{-\alpha_1 t}$ since at $t = 0$, $N_1 = N$. The solution of the second equation is

$$N_2 = A\varepsilon^{-\alpha_1 t} + B\varepsilon^{-\alpha_2 t},$$

where A and B are constants When $t = 0$, $N_2 = 0$, so that $B = -A$ and therefore

$$N_2 = A\left(\varepsilon^{-\alpha_1 t} - \varepsilon^{-\alpha_2 t}\right).$$

Hence
$$\frac{dN_2}{dt} = -\alpha_1 A\varepsilon^{-\alpha_1 t} + \alpha_2 A\varepsilon^{-\alpha_2 t}.$$

When $t = 0$, we have $N_1 = N$ and $N_2 = 0$, so that, when $t = 0$,

$$\frac{dN_2}{dt} = \alpha_1 N = A(\alpha_2 - \alpha_1),$$

and therefore
$$A = \frac{\alpha_1 N}{\alpha_2 - \alpha_1}.$$

Hence
$$N_2 = N\frac{\alpha_1}{\alpha_2 - \alpha_1}\left(\varepsilon^{-\alpha_1 t} - \varepsilon^{-\alpha_2 t}\right).$$

In a similar way we find that

$$N_3 = \frac{N\alpha_1\alpha_2}{(\alpha_3 - \alpha_1)(\alpha_2 - \alpha_1)(\alpha_3 - \alpha_2)}\left\{(\alpha_3 - \alpha_2)\varepsilon^{-\alpha_1 t} - (\alpha_3 - \alpha_1)\varepsilon^{-\alpha_2 t} + (\alpha_2 - \alpha_1)\varepsilon^{-\alpha_3 t}\right\};$$

and similar expressions may be found for N_4, N_5, &c.

The original substance gradually disappears and the final product gradually increases in amount until $N_n = N$. The intermediate products increase up to a maximum amount and then gradually disappear. The time t_2 at which N_2 is a maximum is got by putting $dN_2/dt = 0$, which gives

$$t_2 = \frac{1}{\alpha_2 - \alpha_1}\log\frac{\alpha_2}{\alpha_1}.$$

In the same way, the time t_3 at which N_3 is a maximum is found to be given by

$$(\alpha_2 - \alpha_3)\alpha_1\varepsilon^{-\alpha_1 t_3} + (\alpha_3 - \alpha_1)\alpha_2\varepsilon^{-\alpha_2 t_3} + (\alpha_1 - \alpha_2)\alpha_3\varepsilon^{-\alpha_3 t_3} = 0.$$

An important case is that where α_1 is very small compared with the other α's. In this case N_1 remains nearly constant and equal approximately to N for a long time.

The equation $\frac{dN_2}{dt} = \alpha_1 N_1 - \alpha_2 N_2$, with $\alpha_1 N_1$ constant, gives

$$N_2 = N_1 \frac{\alpha_1}{\alpha_2}(1 - \varepsilon^{-\alpha_2 t}),$$

which shows that N_2 increases to a nearly constant value $N_1 \frac{\alpha_1}{\alpha_2}$.

The equation
$$\frac{dN_3}{dt} = N_2\alpha_2 - \alpha_3 N_3,$$

with N_2 constant and equal to $N_1\alpha_1/\alpha_2$, becomes

$$\frac{dN_3}{dt} = N_1\alpha_1 - \alpha_3 N_3,$$

which gives
$$N_3 = N_1 \frac{\alpha_1}{\alpha_3}(1 - \varepsilon^{-\alpha_3 t}),$$

so that N_3 also increases to a constant value $N_1\alpha_1/\alpha_3$ In the same way we find that N_4 becomes constant and equal to $N_1\alpha_1/\alpha_4$, and so on. Hence in this case we find that eventually

$$N_1\alpha_1 = N_2\alpha_2 = N_3\alpha_3 = N_4\alpha_4 = N_5\alpha_5 = \dots ,$$

provided the α's are large compared with α_1. Since $\alpha_n = 0$, we do not get $N_1\alpha_1 = N_n\alpha_n$, and clearly $\frac{dN_n}{dt} = \alpha_{n-1}N_{n-1} = N_1\alpha_1$, so that the final product increases at the same rate that the first body diminishes, while all the intermediate products remain nearly constant in amount. The amounts of the intermediate products are all proportional to N_1 and so diminish very slowly as N_1 diminishes. When the state for which

$$N_1\alpha_1 = N_2\alpha_2 = N_3\alpha_3 = \dots$$

has been attained, the mixture is said to be in a state of radioactive equilibrium, because each intermediate product is produced almost as fast as it decomposes.

For any intermediate product we have

$$\frac{dN_S}{dt} = \alpha_{S-1}N_{S-1} - \alpha_S N_S,$$

where $S = 2, 3, 4, \dots , n-1$.

It is clear that if N_S is constant for any value of S from 2 to $n-1$, then

$$\alpha_{S-1}N_{S-1} = \alpha_S N_S,$$

so that
$$\alpha_1 N_1 = \alpha_2 N_2 = \alpha_3 N_3 = \alpha_4 N_4 = \dots = \alpha_{n-1}N_{n-1}.$$

Since
$$\frac{dN_n}{dt} = \alpha_{n-1}N_{n-1},$$

we see that in this case $N_n = \alpha_1 N_1 t +$ constant, so that the amount of the final product increases at a uniform rate. We have also $\alpha_1 N_1 =$ constant, so that in order that the intermediate products should remain constant N_1 must be supposed kept constant in some way. This condition is approximately satisfied when α_1 is very small. The product $\alpha_S N_S$ is equal to the number of atoms of the body of number S which decompose in unit time, so that, in a mixture in radioactive equilibrium, the number of atoms decomposing is the same for all the intermediate products.

The number of α-rays emitted by any one of the products present is therefore

the same as the number emitted by any other product which emits α-rays. The same is true of the emission of β-rays by the atomic nuclei The amounts of the different bodies present in the mixture are inversely as their activities, so that the bodies which decompose rapidly are present in small amounts. Uranium minerals are examples of such a mixture of radioactive bodies in a state of radioactive equilibrium.

4. Table of Products formed from Uranium.

The following table gives the products formed from uranium. Each product is formed by the decomposition of the atoms of the preceding product in the table. The first column gives the names and the second the chemical symbols of the products. The third column gives the atomic weights. These are got from that of uranium by subtracting 4 when α-rays are emitted. The next column gives the atomic number. This increases by one when β-rays are emitted and diminishes by two when α-rays are emitted. The atomic weight of radium has been found directly by chemical analysis. The fifth column gives the symbol of an isotope of the product, that is, an element which has the same atomic number and identical chemical properties. The chemical properties of radium are very similar to those of barium, and radon is a chemically inactive gas like argon. Its atomic weight has been determined by measuring its density. The sixth column headed T gives the half-value period. The next column gives the rays emitted. The last column gives the velocity of the rays as a fraction of that of light. Rays in brackets are relatively feeble.

Two other similar series of radioactive products are known, one starting with thorium and ending with lead, and the other of uncertain origin containing actinium.

Name	Symbol	Atomic Weight	Atomic Number	Isotope.	T	Radiation	Velocity of Rays
Uranium I	U-I	238	92	--	$1 \cdot 67 \times 10^9$ yr	α	0·0156
Uranium X_1	U-X_1	234	90	Th	24·6 days	β	—
Uranium X_2	U-X_2	234	91	Pa	1·15 min.	$\beta(\gamma)$	--
Uranium II	U_{II}	234	92	U	2×10^6 yr.	α	0·0179
Ionium	Io	230	90	Th	$6 \cdot 9 \times 10^4$ yr	α	0·0185
Radium	Ra	226	88	—	1690 yr.	$\alpha(\beta, \gamma)$	0·050
Ra Emanation (Radon)	Rn	222	86	--	3·85 days	α	0·054
Radium A	Ra-A	218	84	Po	3·0 min.	α	0·0565
Radium B	Ra-B	214	82	Pb	26·8 min.	$\beta(\gamma)$	0·36; 0·41, 0·63; 0·70; 0·71
Radium C	Ra-C	214	83	Bi	19·5 min.	β	0·786; 0·862; 0·919; 0·957
Radium C′	Ra C′	214	84	Po	10^{-6} sec.	α	0·0641
Radium D	Ra-D	210	82	Pb	16·5 yr.	(β, γ)	0·33, 0·39
Radium E	Ra-E	210	83	Bi	5 days	β	—
Radium F (Polonium)	Ra F	210	84	Po	136 days	$\alpha(\gamma)$	0·0523
Radium Ω (Lead)	Ra-Ω	206	82	Pb	--		-

Radium-C also gives another product to the extent of 0·03 per cent which is produced with the emission of α-rays. This product is called radium-C″, its atomic weight is 210, its atomic number 81; its isotope thallium and its half-value period 1·4 min. It emits only β-rays. Radium apparently emits both α- and β-rays, which is unusual. The β-rays are feeble, and may be due to a few radium atoms decomposing in a different way from that which gives the emanation. If a radium atom emitted an α-ray and also a β-ray the resulting atom should have atomic weight 222 and atomic number 87. An element with atomic number 87 should have chemical properties similar to those of cæsium, whereas the emanation is chemically inert, as it should be if its atomic number is 86.

5. **Properties of Radon.**

It is only possible here to give an account of the investigations which have been made of the properties of a very few of the radioactive bodies. An account will be given of radium emanation or radon, and some of its products. Radon is produced from radium at a practically constant rate, since the half-value period of radium is 1690 years. The radon atoms do not escape to any great extent from solid radium salts since they are produced throughout the volume of the solid and do not diffuse appreciably through it. The radon therefore accumulates in solid radium salts until its rate of decomposition is equal to its rate of production. If the salt is dissolved in water the radon is able to escape, especially if the water is heated under a low pressure.

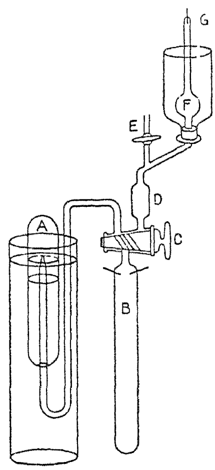


Fig 1

For the purpose of obtaining a supply of radon it is convenient to keep a solution of some radium salt in a glass flask connected to a Toepler pump, by means of which the gases evolved by the solution may be pumped off and collected over mercury. It is found that the solution evolves small amounts of hydrogen and oxygen, which are evidently produced by the action of the radiations on the water. The radon is therefore obtained mixed with oxygen, hydrogen, and water vapour. If all the radon in the solution is collected in this way then after a few days another supply will have accumulated.

The radon may be obtained in the pure state by means of the apparatus shown in fig. 1. The mixture of radon and other gases over mercury in the tube A is let into a bulb B through the stopcock C, the bulb B having been previously exhausted through E, which leads

to a Toepler pump The mixture is exploded in B by means of a spark, and CO_2 is absorbed by means of some caustic potash fused on to the walls of B. The gases are passed into a bulb F through D, which contains phosphorus pentoxide to absorb the water vapour. The bulb F is then surrounded by liquid air and the radon condenses on its walls. The apparatus is next pumped out through E and mercury let in so as to fill it up to the bottom of F. On the liquid air being removed the emanation evaporates and fills F. It can be compressed into a small calibrated capillary tube G at the top of the bulb F by letting in more mercury. In this way the volume of the emanation at different pressures can be measured.

Ramsay and Soddy obtained pure radon from a solution containing 60 mgm. of radium in this way. They found the gas obeyed Boyle's law, and that its volume gradually decreased, falling to half-value in about 4 days. The α-rays penetrate the glass walls and so may not produce gaseous helium in the tube, otherwise the volume should not have diminished. Later experiments have shown that the volume of radon in equilibrium with 1 gm. of radium is approximately 0·6 c. mm. This quantity of radon is called a *curie*.

The number of α-rays emitted by radium, free from other products, has been counted and found to be $3{\cdot}4 \times 10^{10}$ per second from 1 gm. of radium. This should be therefore the number of atoms of radon produced per second. The rate at which the atoms of radon decompose is αN, where $\alpha = 2{\cdot}085 \times 10^{-6}$ and N is the number of atoms of radon. Thus if $3{\cdot}4 \times 10^{10}$ new atoms are produced per second the number existing when equilibrium is reached will be

$$N = \frac{3{\cdot}4 \times 10^{10}}{2{\cdot}085 \times 10^{-6}} = 1{\cdot}63 \times 10^{16}.$$

The number of atoms in 1 c. c. of a monatomic gas at 0° C. and 760 mm. is $2{\cdot}7 \times 10^{19}$, so that the volume of radon in equilibrium with 1 gm. of radium should be

$$\frac{1{\cdot}63 \times 10^{16}}{2{\cdot}7 \times 10^{19}} \times 1000 = 0{\cdot}603 \text{ c. mm.},$$

which agrees very well with the value found experimentally.

The rate of decomposition of radon has been accurately determined by Curie and by Rutherford. Some radon was sealed up in a glass tube, and the intensities of the β and γ radiations coming out of the tube compared at measured times by means of the ionization produced by these rays in a metal ionization chamber. The radon itself emits only α-rays, but the product radium-C gives β- and γ-rays. The half-value periods of radium-A, radium-B, and radium-C are all small

compared with that of radon, so that in a short time the amount of radium-C present in the tube becomes practically proportional to the amount of radon present, and then the rate of decay of the β- and γ-rays is a measure of the decay of the radon. In this way it was found that a for radon is $2{\cdot}085 \times 10^{-6}$ (sec.)$^{-1}$, which gives 3 85 days for the half-value period.

Radon is found to be without chemical properties, like the inert gases helium, neon, argon, krypton, and xenon. Its density was determined by Ramsay and Gray by weighing a fraction of a cubic millimetre of it in a bulb with a microbalance of extraordinary sensibility. The density was nearly 222, taking that of oxygen to be 16, and so agreed with the atomic weight got by deducting four from 226, the atomic weight of radium. Radon therefore comes at the end of the sixth period of the periodic arrangement of the elements, in the same column as the other inert gases. The atomic number of radon is 86, which is got from that of radium 88 by deducting 2 for the emission of one a-ray.

The atomic numbers of the inert gases 2, 10, 18, 36, 54, 86 are equal respectively to 2×1, $2(1+2^2)$, $2(1+2^2+2^2)$, $2(1+2^2+2^2+3^2)$, $2(1+2^2+2^2+3^2+3^2)$, and $2(1+2^2+2^2+3^2+3^2+4^2)$, and so form a regular series with that of radon as the last term.

6. Experimental Study of Products from Radon.

Solid bodies put in contact with radon become coated with a radioactive deposit. This deposit can be concentrated on to a small wire by charging the wire negatively to a few hundred volts. Thus if a solution of radium bromide is put at the bottom of a corked bottle, and a fine wire put through the cork and kept negatively charged, the wire becomes strongly radioactive in a short time. The radiations from the wire may be studied by means of the ionization which they produce in air.

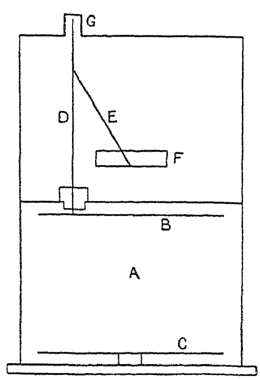


FIG 2

An electroscope suitable for the study of the ionization due to a-rays is shown in fig. 2. It consists of a cubical metal box A containing two parallel metal plates B and C. The plate B is insulated and connected to a rod D which supports a narrow gold leaf E. The gold leaf is in an upper chamber above A. The upper end of D is covered by a removable cap at G which enables the rod to be charged. The deflection of the leaf can be observed with a low-power microscope through the glass window F. A graduated scale in the eyepiece of the microscope enables the deflections to be measured. If a body emitting

α-rays is put on the lower plate C the rays ionize the air and the leaf gradually falls. The number of scale divisions moved by the leaf in a suitable interval of time is proportional to the ionization. To study the β-rays emitted by any radioactive body it is enclosed in a tube or covered with a thin sheet thick enough to stop the α-rays. If it is then put on the plate C the rate of fall of the leaf measures the intensity of the ionization due to the β- and γ-rays. Ionizations due to γ-rays alone may be studied by absorbing the α- and β-rays by means of thick metal plates.

If a bare wire covered with the deposit from radon is put in such an electroscope the ionization is almost entirely due to the α-rays. If the wire has only been exposed to the radon for a short time, say 1 min., the deposit will consist almost entirely of the product of the decomposition of the radon which is called radium-A. It is found that the intensity of the α-rays decreases rapidly for about 10 min., the half-value period being 3 min. It then becomes nearly constant for about 1 hr. and then begins to fall again, with a half-value period of 28 min. The β-ray activity of a wire which has been exposed to radon for less than 1 min. is small initially, but rises in about 20 min. to a maximum and then decreases. This shows that radium-A gives no β-rays, but forms products which give them. The variation of the α-ray activity shows that the products also give some α-rays. On heating the wire to about 600° C. it is found that part of the deposit volatilizes and part remains on the wire. The part remaining on the wire gives α- and β-rays, which both decrease with the half-value period of 19·5 min. The part which evaporates can be condensed on a cool surface. It is inactive at first but then emits α- and β-rays, which increase in intensity to a maximum and then die away.

The results obtained can be explained satisfactorily as follows. The radium-A emits α-rays only and has a half-value period of 3 min. It gives radium-B with a half-value period of 26·8 min., which emits only very feeble β-rays. The radium-B gives radium-C, which emits β-rays and has a half-value period of 19·5 min. This gives radium-C′ with a very short period which gives α-rays. The α-rays from radium-C′ and the β-rays from radium-C both vary with the time in the same way because of the very short period of radium-C′.

It is found that the radioactivity of the deposit does not entirely disappear for a long time, and it appears that several more products are formed. The next one, radium-D, has a half-value period of 16·5 years.

7. Heat evolved by Radium.

When the rays from radioactive bodies are absorbed by matter the energy of the rays is nearly all converted into heat. It is found, for example, that 1 gm. of radium in equilibrium with its products pro-

duces about 120 calories per hour. The total heat produced by a gramme of radium is therefore

$$\frac{120}{60 \times 60} \int_0^\infty \epsilon^{-at}\, dt, \text{ where } a = 1{\cdot}3 \times 10^{-11} \sec^{-1},$$

which is the value of a corresponding to a half period of 1690 years. This is equal to $2{\cdot}57 \times 10^9$ calories. This is enormously greater than the heat evolved per gramme in any chemical reaction. The source of this energy will be discussed in the chapter on Atomic Nuclei.

8. **Artificial Radioactivity.**

M. and Mme Joliot found that aluminium when bombarded by α-rays becomes radioactive and emits positive electrons or positrons. The activity decays exponentially with half-value period 3 min. 15 sec.

It is supposed that the α-rays combine with aluminium nuclei with the emission of a neutron thus:

$${}_{13}\mathrm{Al}^{27} + {}_2\mathrm{He}^4 \rightarrow {}_{15}\mathrm{P}^{30} + {}_0n^1.$$

The ${}_{15}\mathrm{P}^{30}$ is unstable and disintegrates, emitting a positron and a neutrino:

$${}_{15}\mathrm{P}^{30} \rightarrow {}_{14}\mathrm{Si}^{30} + {}_1e^0 + {}_0n^0.$$

The ${}_{14}\mathrm{Si}^{30}$ is a stable isotope of silicon. The aluminium and helium may also react with the emission of a proton:

$${}_{13}\mathrm{Al}^{27} + {}_2\mathrm{He}^4 \rightarrow {}_{14}\mathrm{Si}^{30} + {}_1\mathrm{H}^1.$$

Boron and some other elements give similar effects:

$${}_5\mathrm{B}^{10} + {}_2\mathrm{He}^4 \rightarrow {}_7\mathrm{N}^{13} + {}_0n^1,$$
$${}_7\mathrm{N}^{13} \rightarrow {}_6\mathrm{C}^{13} + {}_1e^0 + {}_0n^0.$$

The half-value period of the ${}_7\mathrm{N}^{13}$ is about 14 min. The positrons from ${}_{15}\mathrm{P}^{30}$ or radio-phosphorus have a continuous energy distribution, like nuclear β-rays, and have an upper energy limit of about 3 *MEV*. The upper limit for the positrons from radio-nitrogen is about 1·5 *MEV*.

Fermi found that most elements are changed into radioactive bodies, emitting β-rays, by neutron bombardment. The neutrons were obtained from a tube containing powdered beryllium and a large quantity of radon. The element to be investigated was made into a thin cylinder which was put round the tube. After exposure to the neutrons for a few minutes the cylinder was removed and tested for β-ray emission with a Geiger counter.

It was found that the β-ray emission was greatly increased when the neutron source and cylinder were surrounded by a large volume of water or paraffin wax. The neutrons collide with the protons in the water or wax, and are slowed down by the collisions so that they have the velocities of gas molecules at the same temperature. These slow neutrons diffuse about in the water or wax, and some of them fall on the metal cylinder. It appears that slow neutrons often combine with atomic nuclei much more readily than the fast neutrons emitted by the source. Some elements are much more efficient in capturing slow neutrons than others. Cadmium, for example, is about 100 times as efficient as many other elements, and gadolinium much more efficient than cadmium. One of the following nuclear reactions occurs when a neutron is captured by the nucleus of an atom ${}_ZA^A$:

$$
\begin{aligned}
{}_ZA^A + {}_0n^1 &\rightarrow {}_ZA^{A+1} & (1)\\
{}_ZA^A + {}_0n^1 &\rightarrow {}_{Z-1}A^A + {}_1\mathrm{H}^1 & (2)\\
{}_ZA^A + {}_0n^1 &\rightarrow {}_{Z-2}A^{A-3} + {}_2\mathrm{He}^4 & (3)
\end{aligned}
$$

The atom ${}_{Z-1}A^A$ produced by (2) usually emits an electron and goes back into ${}_ZA^A$. The electron is emitted with considerable energy which must be supplied by the neutron, so reaction (2) only occurs with fast neutrons. Reaction (1) usually occurs with slow neutrons and is accompanied by the emission of γ-rays. For example, neutrons are absorbed by water and paraffin with the emission of γ-rays and the formation of heavy hydrogen thus:

$$ {}_1\mathrm{H}^1 + {}_0n^1 \rightarrow {}_1\mathrm{H}^2. $$

The nature of a radioactive product can often be determined chemically by dissolving the activated substance and adding to the solution small quantities of elements which might be isotopes of the product. If these elements are then precipitated chemically, one at a time, the product will be precipitated along with its isotope, and its presence in the precipitate can be detected by the rays which it emits. In many cases the product is an isotope of the activated body and so is precipitated with it, showing that the neutron was captured according to reaction (1).

REFERENCES

1. *Radioactivity.* E. Rutherford.
2. *Radiations from Radioactive Substances.* Rutherford, Chadwick and Ellis.

CHAPTER XII

Atomic Nuclei

1. Atomic Numbers and Mass Numbers.

According to Rutherford's theory, an electrically neutral atom consists of a nucleus with a positive charge Ze with Z electrons outside the nucleus. If the atoms in order of increasing atomic weight, beginning with hydrogen, are numbered 1, 2, 3, . . . , then the atomic number, with a few exceptions, is equal to Z.

The nuclear charge Ze can be found by measuring the wave-length of the K_α X-ray line of the atom. Z is then given by Moseley's formula * $\nu_{K_\alpha} = \frac{3}{4}R(Z - a)^2$.

The chemical properties and the spectral line frequencies are determined by Z, and all atoms with the same value of Z are regarded as atoms of the same element.

The atomic weights of atoms, taking that of oxygen equal to 16, determined by the mass spectrograph, are approximately equal to integers A, which are called the atomic weight numbers or mass numbers. The atomic weight of an electron is only 0·00055, so that nearly all the weight of an atom is in its nucleus. The fact that the atomic weights are nearly integers shows that the nuclei are built up out of units with atomic weights about one. Protons and neutrons both have atomic weights about one, so it is supposed that atomic nuclei consist of a number of protons and neutrons closely packed together. The number of protons in a nucleus must be equal to Z, because the charge on a proton is e, and so the number of neutrons N in the nucleus is $A - Z$.

Before the discovery of neutrons it was supposed that a nucleus consisted of A protons and $A - Z$ electrons, but it was difficult to see how an electron could be inside such a small particle as a nucleus. The radius of a nucleus is supposed to be about 15×10^{-13} cm., and the de Broglie wave-length of an electron of mass m moving with velocity v is h/mv. For the electron to be certainly inside the nucleus

* See p. 189.

the wave-length would have to be smaller than the nuclear radius, so that a lower limit for mv is given by $15 \times 10^{-13} = h/mv$, or

$$mv = \frac{6\cdot5 \times 10^{-27}}{15 \times 10^{-13}} = 4 \times 10^{-15}.$$

The rest mass of an electron is only 9×10^{-28}, so that the velocity v must be practically equal to the velocity of light, and so

$$m = \frac{4 \times 10^{-15}}{3 \times 10^{10}} = 1\cdot3 \times 10^{-25},$$

which is about 150 times the rest mass. It is difficult to see how an electron with such enormous energy could be kept inside the nucleus.

The atomic number Z and mass number A of an atom with chemical symbol W may be written with the symbol thus: ${}_ZW^A$. For example, a lithium atom with $Z = 3$ and $A = 6$ may be indicated by ${}_3Li^6$. Since all atoms with the same Z have the same chemical symbol the suffix is often omitted. Table I gives the chemical symbols, atomic numbers, and mass numbers of several atoms.

TABLE I

Element.	Symbol.	Atomic Number.	Mass Numbers.
Hydrogen	H	1	1, 2, 3
Neutron	n	0	1
Helium	He	2	3, 4
Oxygen	O	8	16, 17, 18
Fluorine	F	9	19
Argon	A	18	36, 38, 40
Calcium	Ca	20	40, 42, 43, 44
Tin	Sn	50	112, 114, 115, 116, 117, 118, 119, 120, 122, 124
Tellurium	Te	52	122, 123, 124, 125, 126, 128, 130
Xenon	Xe	54	124, 126, 128, 129, 130, 131, 132, 134, 136
Cæsium	Cs	55	133
Mercury	Hg	80	196, 198, 199, 200, 201, 202, 204
Radium	Ra	88	226
Uranium	U	92	234, 238

Atoms with the same atomic number but different mass numbers are called isotopes, and atoms with the same mass numbers but different atomic numbers are called isobars; for example, there are Sn, Te and Xe atoms with mass numbers 124, and A and Ca atoms with mass numbers 40.

2. Energies of Formation.

The atomic weights are all a little less than the sum of the atomic weights of the Z hydrogen atoms and N neutrons which are combined

in any atom. The atomic weight of hydrogen is 1·0079, and that of a neutron is 1·0083, so that the loss of weight for oxygen ${}_8O^{16}$ is

$$8 \times 1{\cdot}0079 + 8 \times 1{\cdot}0083 - 16 = 0{\cdot}1296,$$

and the loss for mercury ${}_{80}Hg^{200}$ with atomic weight 200·016 is

$$80 \times 1{\cdot}0079 + 120 \times 1{\cdot}0083 - 200{\cdot}016 = 1{\cdot}612.$$

These losses divided by the mass numbers give $0{\cdot}1296/16 = 0{\cdot}0081$ and $1{\cdot}612/200 = 0{\cdot}00806$, which are nearly equal. It is roughly true that the weight losses are proportional to the mass numbers except for the elements with very small atomic weights.

This loss of weight when an atom is formed by the combination of neutrons and hydrogen atoms is attributed to a loss of energy. According to Einstein energy has mass, and the mass of energy E is equal to E/c^2, where c is the velocity of light. The neutrons and protons are supposed to attract each other strongly when very near together, so that in the formation of an atom work is done by these attractions, and some of the energy set free must be supposed to escape as radiation or otherwise. The loss of weight is therefore a measure of the energy of formation or binding energy of the atom.

It is convenient to express the energies of formation in electron volts, that is, in terms of the work done on an electron when it moves across a potential difference of one volt. If we put $Pe = mc^2$, then P is the potential difference required to give an electron the energy in a mass m. If $\mathcal{N}$ is the number of molecules in one mol or molecular weight in grammes, then the faraday F is equal to $\mathcal{N}e$. This gives $P = \mathcal{N}mc^2/F$. If m is the mass of one atom of unit atomic weight, then $\mathcal{N}m = 1$, so that $P = c^2/F$. With P in volts and F in coulombs this gives $P = 10^{-7}c^2/F = 931 \times 10^6$. Hence if MEV denotes one million electron volts, the energy of one unit of atomic weight or unit mass number is 931 MEV.

The energy of formation of oxygen O^{16} is therefore $0{\cdot}1296 \times 931 = 120$ MEV, and that of mercury is $1{\cdot}612 \times 931 = 1500$ MEV. The energy of formation per unit mass number is therefore about 8 MEV for elements with A greater than 16.

The energy of formation in MEV is given in terms of Aston's packing fraction δ by the expression

$$7{\cdot}719A - 0{\cdot}093A\delta - 0{\cdot}372Z,$$

taking ${}_1H^1 = 1{\cdot}0079$ and ${}_0n^1 = 1{\cdot}0083$. The first term is usually much larger than the sum of the other two.

The nucleus may be compared with a small drop of water. The energy required to evaporate the drop is proportional to its mass, so its energy of formation from gaseous water molecules per molecule

is a constant. This is explained by supposing that the molecules attract each other only when very close together, so that each molecule is only appreciably acted on by a small number of other molecules all within a certain very small distance of it. This makes the energy of formation proportional to the number of molecules in the drop. We may therefore suppose that the neutrons and protons in a nucleus attract each other strongly only when extremely close together, so that each particle is only acted on by a very few particles which are next to it.

The protons, of course, repel each other with a force e^2/r^2, which must diminish the energy of formation. This electrostatic effect can be roughly estimated if we assume the protons to be uniformly distributed over the volume, assumed to be spherical, of the nucleus. The potential at the surface of a sphere of electricity of charge density ρ and radius r is $\frac{4}{3}\pi\rho r^3/r$, so the work required to increase its radius by dr is $\frac{4}{3}\pi\rho r^2 4\pi r^2 \rho\, dr$. Integrating this from 0 to r, we get $16\pi^2\rho^2r^5/15$. The density ρ is equal to $Ze/\frac{4}{3}\pi r^3$, so the potential energy is

$$\frac{16\pi^2r^5}{15} \times \frac{9Z^2e^2}{16\pi^2r^6} = \frac{3}{5}\frac{Z^2e^2}{r}.$$

The radius of the radioactive nuclei, for which A is about 222, given by Gamow's theory of α-ray disintegrations is about 15×10^{-13} cm. If r is proportional to $A^{1/3}$ so that $r = r_1A^{1/3}$, where r_1 is the value of r for a nucleus with $A = 1$, we get

$$r_1 = \frac{15 \times 10^{-13}}{222^{1/3}} = 2{\cdot}5 \times 10^{-13}.$$

The electrostatic energy in a nucleus is therefore about $\frac{3}{5}\frac{Z^2e^2}{r_1A^{1/3}}$. Putting $\frac{3}{5}\frac{e^2}{r_1} = Pe$, we get

$$\frac{3 \times 4{\cdot}8 \times 10^{-10} \times 300}{5 \times 2{\cdot}5 \times 10^{-13}} = P \text{ volts},$$

which gives $P = 0{\cdot}35 \times 10^6$. The electrostatic energy is therefore $0{\cdot}35Z^2/A^{1/3}$ *MEV*.

The most stable atoms of low atomic weight contain equal or nearly equal numbers of neutrons and protons. This suggests that the energy of formation of these elements is a maximum for a given A when the number of neutrons N is equal to the number of protons Z.

3. Weizsäcker's Semiempirical Equation.

An expression for the energy of formation E was suggested by Weizsäcker, and after being slightly simplified by Bethe is

$$E = \epsilon A - \alpha(N - Z)^2/A - 0{\cdot}35Z^2/A^{1/3} - \beta A^{2/3}.$$

The second term, when Z is small so that the third term can be neglected, makes E a maximum for a given A when $Z = N$. The last term is proportional to the surface area of the nucleus and is analogous to the surface tension energy of a drop of water.

Differentiating with respect to Z and putting $dE/dZ = 0$, we get, since $N - Z = A - 2Z$,

$$\alpha = 0{\cdot}175 \frac{ZA^{2/3}}{A - 2Z}.$$

Table II gives the values of the constant α given by this equation with the values of A and Z for several elements.

TABLE II

Element.	A	Z.	α.
Hg	200	80	12·0
Sm	150	62	11·8
Mo	100	42	9·9
As	75	33	11·5
Ti	50	22	8·8

The values of α do not vary much, so we take $\alpha = 11$. The constants ϵ and β are found to be 11·2 and 8 respectively by means of the values $E = 1500$ for Hg^{200} and $E = 120$ for O^{16}. The equation giving E in MEV is then

$$E = 11{\cdot}2A - 11(N - Z)^2/A - 0{\cdot}35Z^2/A^{1/3} - 8A^{2/3}.$$

Table III gives several values of E given by this equation and the values got from the atomic weights.

TABLE III

Element.	$E(MEV)$ Calculated.	Found.	E/A Calculated.
${}_{8}O^{16}$	119	120	7·4
${}_{80}Hg^{200}$	1495	1500	7·5
${}_{18}A^{40}$	317	330	7·9
${}_{36}Kr^{82}$	650	688	7·9
${}_{54}Xe^{134}$	1030	1080	7·7
${}_{92}U^{238}$	1737	1630	7·3

It appears that Weizsäcker's equation gives values of E agreeing fairly well with the values got from the atomic weights. The values of E/A are smaller than the value 11·2 given by the first term, so that the electrostatic and surface tension energies, together with the energy proportional to $(N - Z)^2$, are about 30 per cent of the energy $11{\cdot}2A$ due to the attractions between the neutrons and protons.

4. Stability of Atoms.

If an atom ${}_ZA^A$ disintegrates, emitting an α-ray, the kinetic energy of the α-ray is given by

$$\text{K.E.} = {}_2E^4 - ({}_ZE^A - {}_{Z-2}E^{A-4}),$$

where ${}_ZE^A$ denotes the energy of formation of an atom ${}_ZA^A$. For example, the values of ${}_{84}E^{216}$ and ${}_{82}E^{212}$ given by Weizsäcker's formula differ by about 25 MEV, so that since the energy of formation of ${}_2He^4$ is about 27 MEV, we get for the kinetic energy of the α-ray in this case about 2 MEV. This is smaller than the observed value, which is 7 MEV. However, Weizsäcker's formula should not be expected to give the energy of an α-ray exactly, partly because the constants in it are only rather roughly known, and partly because it only gives the general average variation of E with A and Z, so that it should not be expected to give the difference between two nearly equal values of E exactly.

Weizsäcker's formula gives negative values for the kinetic energy of an α-ray for values of A smaller than about 200, which indicates that elements with such values of A cannot emit α-rays, which is in agreement with the facts.

It is found that there are very few isotopes with even A and odd Z. This is shown in the following table.

Numbers of Isotopes

A	Z	N.	Number.
Even	Even	Even	154
Even	Odd	Odd	4
Odd	Even	Odd	55
Odd	Odd	Even	59

The four with even A and odd Z all have $N = Z$ and small atomic weights. They are ${}_1H^2$, ${}_3Li^6$, ${}_5B^{10}$ and ${}_7N^{14}$. With A even and Z even, N and Z are both even, so that all the neutrons can be arranged in pairs and also all the protons. With A even and Z odd, N is also odd, so that the neutrons and protons cannot be arranged in pairs. It is supposed that not more than two neutrons or two protons can be in the same proper state in the nucleus, just as in the electronic system of an atom only two electrons with opposite spins can be in the same proper state. Two similar particles in the same proper state are supposed to be near enough together to interact appreciably, but not if in different states. With these assumptions it follows that when A is even and N odd there will be one proton and one neutron in higher energy levels than the rest, because all the lower levels will be filled with pairs of particles. The energy of formation will therefore be greater with A even and Z even than with A even and Z odd. An

atom with A even and Z odd will therefore emit an electron, a neutron changing into a proton, so becoming an atom with A even and Z even, thus:

$$\bar{e} + {}_ZA^A \rightarrow {}_{Z+1}A^A + \bar{e} + K,$$

where K is the kinetic energy of the emitted electron in atomic weight units The atom will absorb an electron into its electronic system when Z changes to $Z + 1$, so that $K = {}_ZA^A - {}_{Z+1}A^A$. K will be positive because ${}_{Z+1}E^A$ is much greater than ${}_ZE^A$, for the reasons just mentioned. Thus atoms with A even and Z odd are unstable and so are not found in nature. With A odd one or other of N and Z is odd, so that there is always one particle in a higher level, and so atoms with A odd and Z odd have nearly the same energy as with A odd and Z even. We might therefore expect about equal numbers of atoms with A odd, and Z even or odd, as is actually found.

When $N = Z$ an exception occurs, because then with A even and Z and N odd the extra proton and neutron are in approximately equal energy levels and so can interact. In this case, therefore, ${}_{Z+1}E^A$ may not be greater than ${}_ZE^A$, and atoms with A even and Z odd may be stable. N and Z are only equal when A is small, so we only get A even and Z odd with light elements. The above argument involves the assumptions that two neutrons or two protons when close enough together attract each other strongly and that the same thing is true of a neutron and a proton.

5. The Very Light Atoms.

Since two neutrons attract each other strongly when very near together, a dineutron ${}_0A^2$ may be possible. Such dineutrons may exist, but they will be extremely difficult to detect if they do. In the same way a diproton ${}_2A^2$ may be possible, but the electrostatic repulsion between two protons must make the formation of a diproton very unlikely. A third neutron would not be expected to combine with a dineutron, because it would have to be in a higher level than the first two and so could not interact with them, and for the same reason three protons could not combine, and moreover the electrostatic repulsion would make it unlikely even if they could.

A neutron and a proton should combine, and this combination is the deuteron or heavy hydrogen atom ${}_1H^2$. Either a neutron or a proton can combine with ${}_1H^2$, giving ${}_1H^3$ and ${}_2He^3$, because in ${}_1H^2$ there are two vacant levels equal to those occupied. A neutron can combine with ${}_2He^3$ or a proton with ${}_1H^3$, giving ${}_2He^4$ and filling up the vacant equal level. ${}_2He^4$ is the helium atom, which is therefore a saturated body, because another particle must be in a higher level and so cannot combine with it. The energies of formation of ${}_1H^2$,

$_1H^3$, $_2He^3$ and $_2He^4$ are approximately 2, 8, 7 and 28 *MEV* respectively.

A great many attempts have been made to explain these energies of formation by assuming various types of interaction between the particles, and trying to solve the Schrödinger equations so as to get the energy proper values.

6. Theory of the Deuteron.

The Schrödinger equation for the deuteron $_1H^2$ may be written in terms of co-ordinates giving the position of the proton with respect to the neutron, so ignoring the motion of the centre of gravity. If these co-ordinates are x, y, z, then the effective mass for the relative motion is $MM'/(M + M')$, where M is the mass of the proton and M' that of the neutron. Neglecting the small difference between M and M', this is $M/2$, so that the equation is then

$$\frac{h^2}{4\pi^2 M} \Delta w(x) + Ew(x) = V(x)w(x),$$

where x stands for x, y, z, and $V(x)$ is the mutual potential energy. Schrodinger's equation for the deuteron, if we suppose $w(x)$ is a function of r only and put $w(x) = u(r)/r$, becomes

$$\frac{h^2}{4\pi^2 M} \frac{\partial^2 u}{\partial r^2} + Eu = Vu.$$

V and E are both negative, so we change their signs for convenience, making the equation

$$\frac{h^2}{4\pi^2 M} \frac{\partial^2 u}{\partial r^2} + Vu = Eu.$$

If we take V to be zero or very small when $r > a$, then for $r > a$ we have

$$\frac{h^2}{4\pi^2 M} \frac{\partial^2 u}{\partial r^2} = Eu,$$

so that $u_{r>a} = Ae^{-\alpha r}$, where $\alpha = \frac{2\pi}{h}\sqrt{ME}$.

When $r < a$, we may suppose V approximately constant, so that

$$\frac{h^2}{4\pi^2 M} \frac{\partial^2 u}{\partial r^2} = -(V - E)u,$$

and so $u = B \sin \beta r$, where $\beta = \frac{2\pi}{h}\sqrt{M(V-E)}$. At $r = a$ the two solutions must join smoothly so that the two values of $\left(\frac{1}{u}\frac{\partial u}{\partial r}\right)_{r=a}$ must be equal, and so

$$-\alpha = \beta \frac{\cos \beta a}{\sin \beta a} \quad \text{or} \quad \cot \beta a = -\sqrt{E/(V-E)}.$$

When a is very small V/E must be large, so that $\cot \beta a$ must be small and $\beta a = \pi/2$ nearly. We have therefore $\frac{2\pi a}{h}\sqrt{VM} = \frac{\pi}{2}$ or $Va^2 = h^2/16M$ approximately when a is very small. The negative energy E is given by

$$\frac{2\pi a}{h}\sqrt{M(V-E)} = \frac{\pi}{2} + \sqrt{\frac{E}{V-E}}$$

approximately, where V is the negative potential. Since V/E is large, this gives

$$\sqrt{E/V} = \frac{2\pi a}{h}\sqrt{MV} - \pi/2.$$

It appears that if $V = h^2/16Ma^2$ exactly, then $E = 0$, so that V must be greater than $h^2/16Ma^2$ for the proton to be able to combine with the neutron.

The constant $\alpha = \frac{2\pi}{h}\sqrt{ME}$ is the reciprocal of a length which may be regarded as the radius of the deuteron. With $E = 2$ MEV it is about 4×10^{-13} cm. More or less similar results are obtained with other values of $V(r)$, such as $Ae^{-r/a}$ and Ae^{-r^2/a^2}.

7. Collisions between Neutrons and Protons.

When neutrons are passed through hydrogen in a Wilson expansion chamber, there are collisions between the protons and neutrons, and the directions and lengths of the resulting proton tracks can be observed. It is found that a large fraction of the tracks are along directions which nearly coincide with the direction of the incident neutrons.

The wave-length of the de Broglie neutron waves is equal to h/mv, so that for neutrons with velocities about 1/20 the velocity of light, the wave-length is 26×10^{-13} cm., which is much longer than the greatest distance at which the interaction between a neutron and proton is appreciable, which, as we have just seen, is probably less than 4×10^{-13}. We should therefore expect the de Broglie waves of a neutron to be scattered equally in all directions by a proton, relative to a co-ordinate system in which the centre of gravity of the particles

is at rest, because when the scattering volume has dimensions small compared with the wave-length there can be no interference between the scattered waves. The experimental results just mentioned appear therefore to be contrary to the theory that the interaction between a neutron and proton is very small at distances greater than 4×10^{-13} cm.

8. Photoelectric Disintegration of Deuterons.

When the γ-rays from ThC″ are passed through heavy hydrogen in a Wilson expansion chamber, tracks are observed which are believed to be proton tracks due to the γ-ray photons colliding with deuterons and breaking them up into a neutron and proton. The photons have energy 2·62 MEV, and the kinetic energy of the protons, estimated from the length of their tracks, is about 0·25 MEV. The neutrons must have practically equal energy so that the binding energy of the deuteron is

$$2{\cdot}62 - 0{\cdot}50 = 2{\cdot}12\ MEV.$$

The target area for collisions between the photons and deuterons was estimated as about 6×10^{-28} cm.^{2} from the number of tracks observed.

We have seen above * that the chance of radiation producing a transition of an electron in an atom from a state with energy E_n to another with E_m is equal to $\frac{8\pi^3}{ch^2}\left|(ez)_{0mn}\right|^2 E_\nu$, where $(ez)_{0mn}$ is a matrix element of the electric moment ez of the electron along the z-axis, and E_ν is the energy in the radiation per cm.^{2} per unit range of frequency.

If the incident radiation per cm.^{2} is just one photon of energy $h\nu$ in a wave group of length l, then we have $\Delta E \,.\, \Delta t = h$, where ΔE and Δt are the uncertainties in the energy and time of arrival of the photon. This gives $h\Delta\nu\Delta t = h$, so that $\Delta\nu = 1/\Delta t$. The energy E_ν per unit range of frequency in the group is therefore $h\nu/\Delta\nu$ or $h\nu\Delta t$. The chance of a transition due to the group is therefore $\frac{8\pi^3}{ch^2}\left|(ez)_{0mn}\right|^2 h\nu\Delta t$, so that the target area σ of the atom for a transition due to the photon is given by

$$\sigma = \frac{8\pi^3\nu}{ch}\left|(ez)_{0mn}\right|^2 \Delta t.$$

We may use this expression to calculate the target area of a deuteron for disintegration by a γ-ray photon. The proper function of the proton in the deuteron is $w(r) = Ae^{-\alpha r}/r$, where $\alpha = 2\pi\sqrt{ME}/h$ for $r > a$. Since a is very small we may use this value of $w(r)$ in calculating the matrix element. When normalized $w(r)$ is

* See p. 137.

$\left(\frac{a}{2\pi}\right)^{1/2} e^{-ar}/r$. The proper function for the proton after the disintegration will be that of a free particle.

Schrödinger's equation for the free particles with kinetic energy ϵ is

$$(h^2/4\pi^2 M)\Delta U + \epsilon U = 0.$$

If we change to polar co-ordinates then, as with the equation for the hydrogen atom, we find $U = S(\theta, \phi)u(r)/r$ and $u(r)$ is a solution of

$$\frac{h^2}{4\pi^2 M}\left(\frac{d^2 u(r)}{dr^2} - \frac{l(l+1)}{r^2}u(r)\right) + \epsilon u(r) = 0.$$

The selection rule for l requires it to change by ± 1, so, since $l = 0$ for the deuteron, we take $l = 1$ for the free particles and so have

$$\frac{h^2}{4\pi^2 M}\left(\frac{d^2 u}{dr^2} - \frac{2u}{r^2}\right) + \epsilon u = 0.$$

The solution of this is $u = B\left(-\cos kr + \frac{\sin kr}{kr}\right)$, where $k = 2\pi\sqrt{M\epsilon}/h$ and ϵ is the kinetic energy of the particles. Also with $l = 1$,

$$S(\theta, \phi) = (3/4\pi)^{1/2}\cos\theta,$$

so

$$U = \left(\frac{3}{4\pi}\right)^{1/2}\frac{B\cos\theta}{r}\left(-\cos kr + \frac{\sin kr}{kr}\right),$$

which has to be normalized properly.

The incident radiation lasts for the time Δt, so, if the photon causes a transition, the proton will be emitted during this time interval. The proton will therefore be represented by a wave group of radial length $v\Delta t$, where v is the velocity of the proton. We must therefore normalize $u(r)$ for one proton between $r = 0$ and $r = v\Delta t$, so that

$$\int_0^{v\Delta t} u(r)^2\,dr = 1.$$

This gives

$$B^2\int_0^{v\Delta t}\left(\cos^2 kr - 2\frac{\cos kr\sin kr}{kr} + \frac{\sin^2 kr}{k^2r^2}\right)dr = 1,$$

or $B^2 = 2/v\Delta t$ approximately since $kv\Delta t$ is large. We now put $\sin kr/kr - \cos kr$ equal to the real part of $-e^{ikr}(i + kr)/kr$, so that

$$U = -\left(\frac{3}{2\pi v\Delta t}\right)^{1/2}\frac{\cos\theta}{r}\,Re\, e^{ikr}(i + kr)/kr,$$

where Re means that only the real part is to be taken. The matrix

element $(z)_{0mn}$ is equal to $\frac{1}{2}\int \bar{\psi}(r)\, r\cos\theta U\, d\tau$, where $d\tau = 2\pi r \sin\theta\, d\theta\, r\, dr$ and $\frac{1}{2}r\cos\theta = z$. We have therefore

$$(z)_{0mn} = -\pi\left(\frac{a}{2\pi}\right)^{1/2}\left(\frac{3}{2\pi v\Delta t}\right)^{1/2}\frac{1}{k}\int e^{-ar}\sin\theta\cos^2\theta\, Re\, e^{ikr}(i+kr)\,d\theta\,dr,$$

so

$$(z)_{0mn} = -\frac{1}{k}\left(\frac{a}{3v\Delta t}\right)^{1/2}\int Re\, e^{-r(a-ik)}(i+kr)\,dr.$$

Also

$$\int_0^\infty e^{-r(a-ik)}(i+kr)\,dr = \frac{i}{a-ik}+\frac{k}{(a-ik)^2},$$

and the real part of this is $-2k^3/(a^2+k^2)^2$, so, putting $v = kh/2\pi M$, we get

$$(z)_{0mn} = +\frac{2}{3}\left(\frac{6\pi a M}{h\Delta t}\right)^{1/2}\frac{k^{3/2}}{(a^2+k^2)^2}.$$

This gives

$$\sigma = \frac{64\pi^4 \nu M e^2 a k^3}{3ch^2(a^2+k^2)^4}.$$

Now $h\nu = E+\epsilon$, $ME = h^2a^2/4\pi^2$, and $M\epsilon = h^2k^2/4\pi^2$, so that

$$\sigma = \frac{4e^2hE^{1/2}\epsilon^{3/2}}{3cM(E+\epsilon)^3}.$$

With $E = 2{\cdot}12\ MEV$ and $E+\epsilon = 2{\cdot}62\ MEV$, this gives

$$\sigma = 7\times 10^{-28}\ \text{cm.}^2,$$

which agrees nearly with the value 6×10^{-28} cm.^{2} found experimentally. The area 7×10^{-28} is equal to that of a circle of radius $1{\cdot}5\times 10^{-14}$ cm., which is much smaller than 4×10^{-13} cm., the supposed radius of the deuteron.

9. Nuclear Reactions.

In 1919 Rutherford made the very interesting discovery that when high-velocity α-rays are passed through certain substances rays are produced which have a greater range than the α-rays. These rays were found to be protons. It appears that the alpha ray enters a nucleus and combines with it, and the nucleus immediately disintegrates emitting a proton. For example, with aluminium

$${}_{13}\mathrm{Al}^{27} + {}_{2}\mathrm{He}^{4} \rightarrow {}_{15}A^{31} \rightarrow {}_{14}A^{30} + {}_{1}\mathrm{H}^{1} + E.$$

The atom ${}_{15}A^{31}$ is an isotope of phosphorus, and ${}_{14}A^{30}$ is an isotope of silicon. E is the energy set free in atomic weight units. Such processes

are called nuclear reactions, and a great many have been found to occur since Rutherford's original discovery. In an equation for a nuclear reaction the sum of the superfixes A representing the mass numbers and the sum of the suffixes representing the atomic numbers must be the same on both sides of the equation. Also the sum of the atomic weights on the left must equal the sum of the atomic weights plus the energy E in atomic weight units on the right.

Nuclear reactions in which an electron or positron is emitted are exceptions to these rules, because when an electron is emitted a neutron changes into a proton, and when a positron is emitted a proton changes into a neutron. Thus

$$_{-1}e^0 + {}_ZA^A = {}_{Z+1}A^A + {}_{-1}e^0 + E$$

and

$$_ZA^A = {}_{Z-1}A^A + {}_{+1}e^0 + {}_{-1}e^0 + E.$$

In the first equation an electron $_{-1}e^0$ is emitted, so that Z changes to $Z+1$ and an electron is added to the electronic system, which is indicated by the electron on the left. The energy E is therefore given by $E = {}_ZA^A - {}_{Z+1}A^A$, the two electron masses cancelling. In the second equation a positron is emitted and Z changes from Z to $Z-1$, so that an electron is lost by the electronic systems. The energy E is then given by $E = {}_ZA^A - {}_{Z-1}A^A - 2m$, where m is the rest mass of an electron or positron which is 0·00055 in atomic weight units.

10. Disintegrations due to α-rays.

The disintegration of several elements by α-rays has been investigated very fully by Chadwick and Constable and others. We shall here consider the case of aluminium. The apparatus used is shown in fig. 1. P is a metal plate coated with polonium or radium C′, which emits α-rays. A is a thin aluminium foil on which some of the α-rays fall. The energy of the α-rays striking the foil at A can be varied by altering the pressure of the gas between P and A. The protons emitted at A are detected by the counter BE. This consists of a double metal chamber with an opening at B covered with thin foil and an insulated electrode E. The gas pressure in the chamber is a few millimetres of mercury, and a potential difference is maintained between B and E so that when a proton enters and ionizes the gas more ions are produced by collisions, so that an appreciable charge is received by E which alters its potential. E is connected to an amplifier and oscillograph the deflections of which are recorded, so that the number of protons entering the chamber per second can be determined. Thin sheets of mica can be put between A and B so as to stop protons not having enough energy to penetrate the mica. The α-rays from P did not have enough penetrating power to get into B, so only protons were counted.

If the aluminium foil A is very thin, so that it does not appreciably alter the energy of an α-ray, then all the collisions in it will be with α-rays of the same energy. With such a very thin foil it is found that the number of protons obtained varies periodically with the energy of the α-rays, so that the number of protons has maximum values for a series of α-ray energies equal to 6·6, 5·8, 5·3, 4·9, 4·5, and 4·0 *MEV*.

When the α-ray enters an aluminium atom we get a phosphorus atom ${}_{15}P^{31}$, so that it appears that this atom can exist in a series of states with energies differing by amounts equal to the differences between the α-ray energies which give maximum numbers of protons.

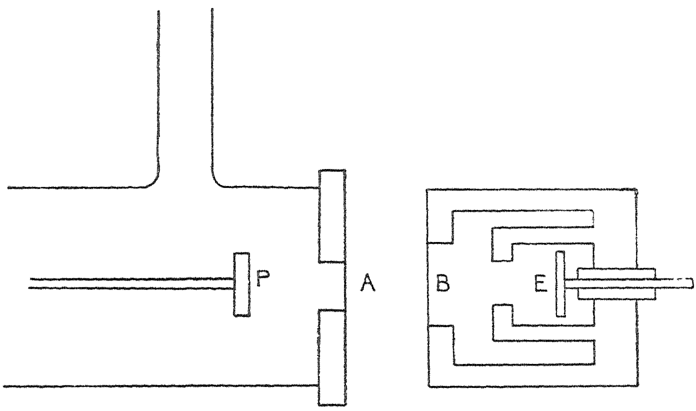


Fig. 1

When the α-ray energy is equal or nearly equal to the energy difference between the Al atom and one of the states of the ${}_{15}P^{31}$ atom, there is a sort of resonance effect, and the chance of the α-ray entering the Al atom is much greater than when this is not the case. The ${}_{15}P^{31}$ atom is unstable and disintegrates, emitting a proton and becoming a silicon atom ${}_{14}Si^{30}$. It is found that for each state of the P atom protons are emitted with two different energies, so it is supposed that the silicon atom may be formed in two different states, one with more energy than the other. For example, the protons due to α-rays with energy 5·3 *MEV* have ranges of 34 and 66 cm. in air at atmospheric pressure, and those due to 4·0 *MEV* α-rays have ranges of 22 and 49 cm. If the silicon atom is formed in an excited state with more than its normal energy, then the proton emitted has less energy and the excited silicon atom must change to its normal state, after emitting the proton, probably with emission of a γ-ray.

Similar results were obtained with fluorine and boron.

11. Disintegrations due to Protons and Deuterons.

The natural radioactive elements like radium do not emit protons or deuterons, so in order to study disintegrations due to high-velocity protons and deuterons it is necessary to obtain them by artificial means.

In order that a proton may enter a nucleus it might be supposed that it would be necessary for the proton to have enough kinetic energy to overcome the repulsion due to the nuclear charge. The electrostatic potential energy of a proton at the surface of a nucleus of mass number A and charge Ze is about $Ze^2/r_1A^{1/3}$, where $r_1 = 2{\cdot}5 \times 10^{-13}$, which is equal to $0{\cdot}6Z/A^{1/3}$ *MEV* approximately. For ${}_3\mathrm{Li}^6$ this is 1 *MEV*, and for ${}_{13}\mathrm{Al}^{27}$ it is 2·6 *MEV*. To obtain protons with such high energies would be difficult, but according to the wave mechanics theory a particle has a finite chance of passing through a narrow region in which its classical kinetic energy is negative, so it seemed possible that protons with only a few hundred thousand volts energy might be able to enter nuclei of low atomic number.

Cockcroft and Walton, working in the Cavendish laboratory, therefore decided to attempt nuclear disintegrations with protons of comparatively low energy, and they constructed an apparatus capable of giving about 900,000 volts. With this apparatus they soon discovered that protons with only 100,000 volts energy could disintegrate several light elements.

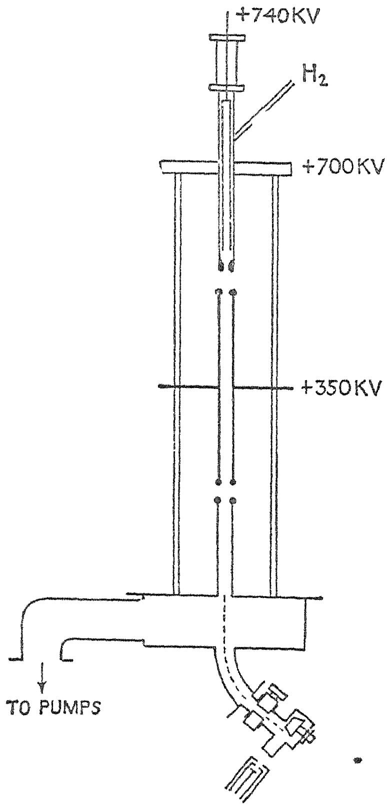


Fig. 2

The latest form of apparatus used by Cockcroft and Walton is shown in fig. 2. It consists of a large vacuum tube made out of two glass cylinders each about 90 cm. long and 30 cm. in diameter. The upper end is closed by a metal plate which carries a hollow electrode consisting of two concentric tubes. A slow stream of hydrogen is admitted into the space between the two tubes, and a discharge can be passed through the hydrogen from the inner tube to the outer one.

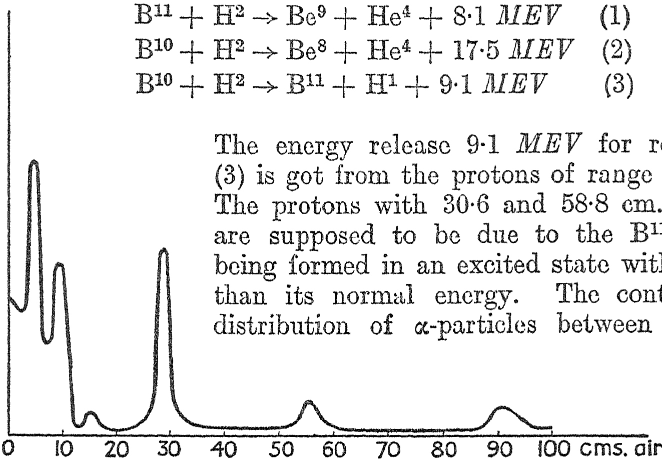
Some of the positive hydrogen ions or protons escape through a small hole at the lower end of the electrode. The upper electrode can be kept at 700,000 volts, and the protons move down the axis of the glass cylinders. A good vacuum is maintained in the apparatus by means of large oil diffusion pumps. A metal plate between the two glass cylinders supports a steel tube as shown, and a similar tube is supported by the plate closing the lower end. The protons are accelerated in the two gaps, and the gaps are designed to prevent spreading of the proton beam, which is about 1 cm. across at the lower plate. The proton beam is deflected by a magnetic field perpendicular to the plane of the paper, and after passing through a wide stopcock strikes a target mounted on a cone. Several targets can be put on the cone, which can be rotated so as to bring any one of them into the proton beam. A window of thin foil is placed about 4 cm. from the target in a direction at right angles to the proton beam. High-velocity atoms emitted by the target pass through the window and are counted with a counter similar to that used by Chadwick and Constable in their experiments with α-rays. The magnetic field can be adjusted so as to deflect the protons, but not other ions of different mass or energy, on to the target. If heavy hydrogen is used instead of ordinary hydrogen, then deuterons ${}_1H^2$ are obtained instead of protons, and the magnetic field can be adjusted to deflect the deuterons on to the target and so separate them from any protons or other ions present due to impurities in the gas.

Thin sheets of mica can be put between the window and the counter, and so the range of the high-speed atoms can be found. An α-ray very near the end of its range gives much more ionization in the counter than when not near the end of its range, and so much larger oscillograph deflections. The number of large deflections per minute is equal to the number of α-rays entering the counter almost at the end of their range. If the target emits α-rays with a definite energy, then the number of large deflections will rise to a sharp maximum, as the thickness of the mica is increased by small steps, when the thickness of the mica is such that the rays enter the counter very near the end of their range. The range of the rays emitted can thus be found, and from the range the kinetic energy. If the target emits protons, the range can be found in the same way. The deflections due to protons near the end of their range are smaller than those due to α-rays, so it is possible to distinguish between α-rays and protons. Fig. 3 shows how the number of large deflections varied with the thickness of mica with a target of fused B_2O_3 on copper bombarded with deuterons of 550 KV energy. The horizontal coordinate is the thickness of air at the atmospheric pressure equivalent to the air and mica between the target and counter.

There are particles emitted by the target with ranges of 4·6, 14·8,

30·6, 58·8, and 91 cm. The particles with ranges 4·6 and 14·8 cm. are α-particles, and the other three are protons. The energies of the emitted particles can be got from their ranges, and the energy released in the reactions can be calculated from the energy of the incident particles and that of the emitted particles assuming momentum to be conserved in the collisions between the incident particles and the atoms in the target. The α-rays and protons are supposed to be due to the reactions

$$B^{11} + H^{2} \rightarrow Be^{9} + He^{4} + 8{\cdot}1\ MEV \quad (1)$$
$$B^{10} + H^{2} \rightarrow Be^{8} + He^{4} + 17{\cdot}5\ MEV \quad (2)$$
$$B^{10} + H^{2} \rightarrow B^{11} + H^{1} + 9{\cdot}1\ MEV \quad (3)$$



The energy release 9·1 *MEV* for reaction (3) is got from the protons of range 91 cm. The protons with 30·6 and 58·8 cm. range are supposed to be due to the B^{11} atom being formed in an excited state with more than its normal energy. The continuous distribution of α-particles between ranges

Fig 3

4·5 and 13 cm. shown in fig. 3 is supposed to be due to the reaction

$$B^{10} + H^{2} \rightarrow 3He^{4} + E.$$

When the disintegrating atom breaks up into more than two particles, a continuous energy distribution results although the total energy is constant, because conservation of momentum is possible with different divisions of the energy among the particles.

Nuclear reactions have been studied with apparatus more or less similar to Cockcroft and Walton's in several laboratories, and a great many reactions have been discovered and their energy releases measured.

12. The Cyclotron.

A very ingenious and valuable method of obtaining protons or deuterons with energies up to 10 *MEV* or more without the use of very high potential differences was devised by Lawrence at the University of California. His apparatus is called a cyclotron.

A particle of mass m and charge e moving with velocity v, in a plane perpendicular to a uniform magnetic field H, describes a circle of radius r given by $mv^2/r = Hev$ so that $mv = Her$. The time T to go once round the circle is given by $T = \frac{2\pi r}{v} = \frac{2\pi m}{He}$ and so is independent of the radius r.

Lawrence's apparatus is shown in fig. 4. It consists of two semicircular metal boxes A and B, with their adjacent sides open, which are put between the poles of a very large electromagnet. These boxes are connected to an electric oscillator so that an alternating potential

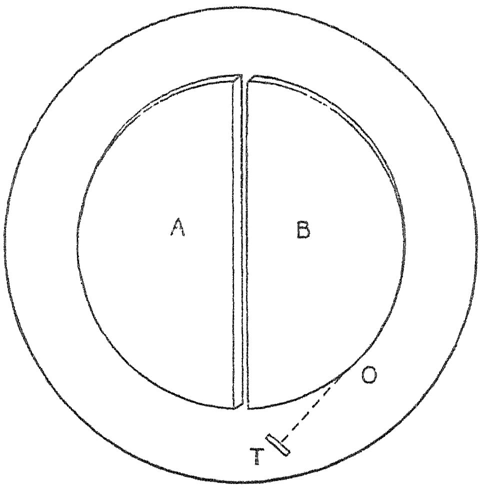


Fig. 4

difference $V_0 \sin pt$ is maintained between them. Positive ions are generated just outside the boxes near the centre, and some of these ions are attracted into a box when it is negative and describe a semicircular path in it. After describing a semicircle the ions pass across into the other box, and are again accelerated and describe another semicircular path, and are again accelerated, and so on. The frequency of the oscillator is adjusted so that the potential difference between the two boxes is reversed while the ions describe a semicircle which requires that $pT = 2\pi$ or $p = He/m$ exactly. With $V_0 = 20{,}000$ volts the energy of the ions is 4 *MEV* after making 100 complete revolutions in the boxes. As the velocity of the ions increases the radius of the path increases proportionally since $r = mv/He$. The kinetic energy is given by $\frac{1}{2}mv^2 = H^2e^2r^2/2m$. For protons with 4 *MEV* energy and $H = 20{,}000$, r is about 14 cm., and with 10 *MEV* energy $r = 23$ cm. To obtain a uniform magnetic field of 20,000 gauss over an area 50 cm. in diameter requires a very large magnet, and magnets weighing up

to 100 tons are used. The semicircular boxes are enclosed in a larger box in which a vacuum is maintained, and the beam of positive ions is allowed to pass through an aperture O in one of the semicircular boxes and to fall on a target T coated with the substance to be investigated. The rays emitted by the target are let out through a window and counted as in Cockcroft and Walton's experiments.

The energies of reactions in which a neutron is emitted cannot be found by the methods so far described, because neutrons do not ionize gases like protons and α-rays. The energy of neutrons can be found by passing them through a Wilson expansion chamber containing hydrogen and observing the lengths of the proton tracks due to collisions between the neutrons and protons. In a head-on collision the neutron gives practically all its kinetic energy to the proton, so that the length of the tracks in the same direction as the incident neutrons gives the energy of the neutrons. Many neutron energies have been measured in this way by Bonner and Brubaker and others. The Wilson expansion chamber has also often been used to verify the nature of nuclear reactions. For example, it has been shown that when a lithium target is bombarded with protons then α-ray tracks occur in pairs, each pair consisting of two equal tracks going in opposite directions from a point on the target. This is what we should expect with the reaction

$$\mathrm{Li}^7 + \mathrm{H}^1 \rightarrow 2\mathrm{He}^4.$$

In the same way it has been shown that when boron is bombarded with protons, with the target in an expansion chamber, then three tracks may be observed all in the same plane and starting from the same point on the target. This verifies the reaction $\mathrm{B}^{11} + \mathrm{H}^1 \rightarrow 3\mathrm{He}^4$. By measuring the lengths of the tracks Dee and Gilbert found the energy release to be 8·7 *MEV*. It was found that usually two of the three tracks are nearly in opposite directions and the third one is much shorter than the other two.

Soon after Rutherford discovered α-ray disintegrations, Chadwick photographed a great many α-ray tracks in air and got several photographs showing the disintegration of a nitrogen atom. The photographs showed a thin proton track and a short thick track starting from the end of an α-ray track. It was supposed that the α-ray enters the nitrogen atom, which then disintegrates emitting a proton. The short thick track is attributed to the ${}_8\mathrm{O}^{17}$ atom formed by the reaction ${}_7\mathrm{N}^{14} + {}_2\mathrm{He}^4 \rightarrow {}_8\mathrm{O}^{17} + {}_1\mathrm{H}^1$. One of these photographs is reproduced at p. 149.

The following table gives a number of typical nuclear reactions and their energy releases. The energies in atomic weight units ($1 = 931$ *MEV*) are nearly all approximately multiples of $q = 0{\cdot}387$

MEV or 0·000415 mass units, and so may be conveniently expressed in terms of this quantity as unit.

Nuclear Reactions

$$
\begin{array}{lll}
{}_6C^{12} + {}_1H^2 \rightarrow {}_6C^{13} + {}_1H^1 + 7q & & (1)\\
{}_6C^{12} + {}_1H^2 \rightarrow {}_7N^{13} + {}_0n^1 - q & & (2)\\
{}_7N^{13} & \rightarrow {}_6C^{13} + e^+ + e^- + 1{\cdot}25\ MEV & (3)\\
& e^+ + e^- + 1{\cdot}25\ MEV = 6q & \\
Li^6 + H^1 & \rightarrow He^4 + He^3 + 9q & (4)\\
H^2 + H^2 & \rightarrow He^3 + n^1 + 7q & (5)\\
Li^6 + H^2 & \rightarrow 2He^4 + 57q & (6)\\
B^{11} + H^1 & \rightarrow 3He^4 + 22q & (7)\\
B^{11} + H^1 & \rightarrow Be^8 + He^4 + 21q & (8)\\
Be^9 + H^2 & \rightarrow Li^7 + He^4 + 19q & (9)\\
B^{11} + H^2 & \rightarrow C^{12} + n^1 + 35q & (10)\\
N^{14} + H^2 & \rightarrow C^{12} + He^4 + 31q & (11)\\
O^{16} + H^2 & \rightarrow N^{14} + He^4 + 8q & (12)\\
n^1 + H^1 & \rightarrow {}_1H^2 + 6q & (13)\\
Li^6 + H^2 & \rightarrow Li^7 + H^1 + 13q & (14)\\
H^2 + H^2 & \rightarrow H^3 + H^1 + 10q & (15)\\
Be^9 + H^2 & \rightarrow B^{10} + n^1 + 11q & (16)\\
Be^9 + H^2 & \rightarrow Be^{10} + H^1 + 12q & (17)
\end{array}
$$

13. Energies of Formation.

Any atom ${}_ZA^A$ may be supposed formed by the hypothetical nuclear reaction

$$ZH^1 + (A - Z)n^1 \rightarrow {}_ZA^A + {}_ZE^A,$$

where ${}_ZE^A$ is the energy of formation of the atom out of protons and neutrons. The energy of any nuclear reaction can be calculated if the energies of formation of the atoms involved are known. For example, the energy E of the reaction $Be^9 + H^2 \rightarrow Li^7 + He^4$ is given by the equation ${}_4E^9 + {}_1E^2 = {}_3E^7 + {}_2E^4 - E$. If we have N different atoms, including ${}_0n^1$, ${}_1H^1$ and ${}_8O^{16}$, then there are $N - 2$ energies of formation. There cannot then be more than $N - 2$ independent nuclear reactions involving only the N elements. For example, the reactions $Li^7 + H^1 \rightarrow 2He^4$, $Li^7 + H^2 \rightarrow 2He^4 + n^1$, and $H^1 + n^1 \rightarrow H^2$ are not independent because the third can be obtained by subtracting the first from the second one.

14. Calculation of Atomic Weights.

If we take ${}_8O^{16} = 16$, then there are $N - 1$ atomic weights remaining, so that, since there are not more than $N - 2$ independent reaction equations, there are not enough to enable the $N - 1$ atomic weights to be calculated. To calculate the atomic weights from the reaction energies it is therefore necessary to assume a value for one other atomic weight besides that of ${}_8O^{16}$. The atomic weight of the electron or positron 0·00055 is very exactly known, and so may be used as the second assumed atomic weight provided one or more of the reaction equations of known energy involves the emission of an electron or positron. The energies of formation may be calculated from the reaction energies without assuming the values of any atomic weights since there are as many independent equations as energies of formation. It is convenient to first calculate the energies of formation and then use them to get the atomic weights.

Equation (13) gives $6q$ for the energy of formation of H^2. Equation (15) then gives $12q = {}_1E^3 - 10q$, so that the energy of formation of ${}_1H^3$ is $22q$. Equation (5) gives for the energy of formation of He^3, ${}_2E^3 = 12q + 7q = 19q$. The equations (4) and (6) give $He^4 + H^1 = He^3 + H^2 - 48q$, so that ${}_2E^4 = (19 + 6 + 48)q$ or the energy of formation of ${}_2He^4$ is $73q$.

Equation (6) gives $146q = {}_3E^6 + 6q + 57q$, so that the energy of formation of Li^6 is $83q$. Equation (14) gives $83q + 6q = {}_3E^7 - 13q$, so that the energy of formation of ${}_3Li^7$ is $102q$. Equation (9) gives ${}_4E^9 = 150q$. Equation (7) gives ${}_5E^{11} = 197q$, and (8) gives ${}_4E^8 = 145q$. Equation (17) gives ${}_4E^{10} = 168q$, and (16) gives ${}_5E^{10} = 167q$. Equation (10) gives ${}_6E^{12} = 238q$, (11) gives ${}_7E^{14} = 271q$, and (12) gives ${}_8E^{16} = 330q$. Equation (1) gives ${}_6E^{13} = 251q$, and (2) gives ${}_7E^{13} = 243q$.

The equation $ZH^1 + (A - Z)n^1 = {}_ZA^A + {}_ZE^A$ gives the atomic weights ${}_ZA^A$ in terms of H^1, n^1 and ${}_ZE^A$. To get the atomic weights from the energies of formation we must therefore first calculate the atomic weights of H^1 and n^1. The equations $7H^1 + 6n^1 = {}_7N^{13} + {}_7E^{13}$ and $6H^1 + 7n^1 = {}_6C^{13} + {}_6E^{13}$ give $n^1 - H^1 = {}_6C^{13} - {}_7N^{13} + {}_6E^{13} - {}_7E^{13}$. Equation (3) gives ${}_7N^{13} - {}_6C^{13} = 6q$ and ${}_6E^{13} - {}_7E^{13} = 251q - 243q = 8q$, so that $n^1 - H^1 = 2q$.* The energy of the hypothetical reaction ${}_8O^{16} = 8\,{}_1H^2 + E$ is given by $330q = 8 \times 6q - E$, so that $E = -282q$. Putting ${}_8O^{16} = 16$, we have therefore ${}_1H^2 = 2 + \frac{282}{8}q = 2 + 35\frac{1}{4}q$. But $n^1 + H^1 = H^2 + 6q$, so that $n^1 + H^1 = 2 + 41\frac{1}{4}q$, and we have $n^1 - H^1 = 2q$. These equations give

$$n^1 = 1 + \tfrac{1}{2}(43\tfrac{1}{4})q,\ H^1 = 1 + \tfrac{1}{2}(39\tfrac{1}{4})q,$$

which with $q = 0{\cdot}000415$ give $n^1 = 1{\cdot}0089745$ and $H^1 = 1{\cdot}0081445$.

* It is possible that ${}_7N^{13} - {}_6C^{13}$ is really equal to $7q$, and that $n^1 + H^1 = {}_1H^2 + 5q$. If so, $n^1 = 1{\cdot}008567$ and $H^1 = 1{\cdot}00815$.

The other atomic weights can now be calculated by means of the equation $ZH^1 + (A - Z)n^1 = {}_ZA^A + {}_ZE^A$. The atomic weights, so calculated, and the energies of formation are given in the following table.

Atomic Weights calculated from Nuclear Reaction Energies

Element.	Atomic Weight.	Energy.
n^1	1·00898	—
H^1	1·00815	—
H^2	2·01463	$6q$
H^3	3·01696	$22q$
He^3	3·01738	$19q$
He^4	4·00394	$73q$
Li^6	6·01691	$83q$
Li^7	7·01800	$102q$
Be^8	8·00830	$145q$
Be^9	9·01520	$150q$
Be^{10}	10·01671	$168q$
B^{10}	10·01629	$167q$
B^{11}	11·01281	$197q$
C^{12}	12·00391	$238q$
C^{13}	13·00752	$251q$
N^{13}	13·01001	$243q$
N^{14}	14·00737	$271q$
O^{16}	16·00000	$330q$

These atomic weights agree very closely with the values obtained by Aston, Bainbridge and others with mass spectrographs.* This agreement of the values obtained by these two quite independent methods is very striking, and confirms the assumption that energy E has mass E/c^2.

The energy of any nuclear reaction can be calculated when the atomic weights of the elements involved are known. The above atomic weights give nuclear reaction energies agreeing with those observed, in most cases, but not in all. For example, they give for the energy of $B^{10} + n^1 \rightarrow Li^7 + He^4$ the value 3·1 *MEV*, whereas the observed value is said to be only 2 *MEV*. Such discrepancies will no doubt be removed when the reaction energies have been more exactly determined.

15. Constitution of Nuclei.

The forces between nuclear particles have a very short range and become saturated—which means that one particle cannot interact strongly with more than a small number of other particles near to it. Leaving out the electrostatic repulsion between the protons, the forces of attraction between a proton and a neutron are believed to be about equal to the attraction between two neutrons or two protons.

The helium nucleus consisting of two protons and two neutrons is

* See p. 236.

believed to be saturated, so that apparently one nuclear particle can not interact strongly with more than three other particles.

On the liquid-drop model of a nucleus it is supposed that a particle in the interior interacts with the twelve particles next to it but that a particle at the surface only interacts with about six nearby particles. The potential energy of a particle therefore gets much smaller when it moves from the surface to the interior. This produces surface tension which prevents the electrostatic repulsion from making the drop get larger.

If we suppose that each particle can only interact with three nearby particles there will be little or no surface tension, so that the electrostatic repulsions between the protons will cause the drop to get as large as possible with each particle interacting with three nearby particles. This means that the drop would change into a hollow spherical shell consisting of a layer one particle thick. Each particle in such a shell would be near three other particles with which it could interact strongly as in the helium nucleus.

This spherical-shell model of a nucleus has been proposed by the writer as a possible alternative to the liquid-drop model.

It is found experimentally that many nuclei have a set of nearly equally spaced energy levels with separations of 0·387 *MEV*. The liquid-drop model gives levels much nearer together than 0·387, with separations getting rapidly smaller as the energy increases. The liquid-drop model does not offer any explanation of the equally spaced levels observed.

The spherical-shell model gives equally spaced levels and with the experimentally found spacing (0·387 *MEV*) enables the nuclear radii to be calculated. The nuclear radii so calculated agree nearly with the radii found by other methods. It is possible therefore that the shell model may prove to be nearer to the truth than the liquid-drop model.

To calculate the frequencies of vibration of the spherical shell, we may suppose it distorted so that its radius r is equal to $a(1+\alpha P_n(\cos\theta))$ where a is the undistorted radius, α a small number and $P_n(\cos\theta)$ a zonal harmonic of order n. θ is the angle between the radius r and a fixed axis through the centre. We have $P_0 = 1$, $P_1 = \cos\theta$, $P_2 = \frac{1}{2}(3\cos^2\theta - 1)$, &c.

Neglecting powers of α greater than two, the volume V of the distorted shell is equal to $\frac{4}{3}\pi a^3\left(1 + \frac{3\alpha^2}{2n+1}\right)$ and the area S to $4\pi a^2\left(1 + \frac{n^2+n+2}{2(2n+1)}\alpha^2\right)$. The potential energy P of the distorted shell, for very small values of α, will be $T\delta S - \frac{2T}{a}\delta V$ where T is the tension in the shell. If the electric charge per unit area on the shell is σ then the force per unit area is $2\pi\sigma^2 = 2T/a$.

The potential energy is therefore given by

$$P = 4\pi a^2 T \frac{n^2+n+2}{2(2n+1)}\alpha^2 - \frac{4}{3}\pi a^3 \frac{2T}{a}\frac{3\alpha^2}{2n+1},$$

or

$$P = 2\pi a^2 T \frac{(n-1)(n+2)}{2n+1}\alpha^2.$$

The kinetic energy of the shell when vibrating can be calculated as follows.

For a nucleus with mass number A the mass per unit area is $Am/4\pi a^2$ where m is the mass of one nuclear particle. The radial velocity, since $r = a(1+\alpha P_n)$, is $aP_n\dot{\alpha}$, so the kinetic energy per unit area is $\frac{1}{2}\frac{Am}{4\pi a^2}a^2 P_n{}^2\dot{\alpha}^2 = \frac{1}{8\pi}AmP_n{}^2\dot{\alpha}^2$. The kinetic energy is therefore equal to $\frac{Am\dot{\alpha}^2}{8\pi}\int_0^\pi P_n{}^2 2\pi a^2 \sin\theta d\theta$, or $\frac{Ama^2\dot{\alpha}^2}{2(2n+1)}$, since $\int_0^\pi P_n{}^2 \sin\theta d\theta = \frac{2}{2n+1}$.

The frequency of oscillation of a simple harmonic oscillator with potential energy $\frac{1}{2}\mu x^2$ and kinetic energy $\frac{1}{2}m\dot{x}^2$ is equal to $\frac{1}{2\pi}\sqrt{(\mu/m)}$ so that the frequency ν_n of the spherical shell is given by

$$\nu_n = \frac{1}{2\pi}\sqrt{\left\{\frac{2\pi a^2 T(n-1)(n+2)}{(2n+1)Ama^2}2(2n+1)\right\}},$$

or

$$\nu_n = \sqrt{\left\{\frac{T(n-1)(n+2)}{\pi Am}\right\}}.$$

The electric field F at the surface of the nucleus is equal to Ze/a^2 so that the force per unit area is $F^2/8\pi = Z^2e^2/8\pi a^4$. The tension T is given by $2T/a = Z^2e^2/8\pi a^4$, so that $T = Z^2e^2/16\pi a^3$. The frequency ν_n is therefore given by

$$\nu_n = \frac{Ze}{4\pi a}\sqrt{\left\{\frac{(n-1)(n+2)}{Ama}\right\}}.$$

This gives the frequencies for very small amplitudes. It is found that the frequency is nearly independent of the amplitude, at any rate when A is large, so the variation of the frequency with the amplitude will be ignored.

Let $\Delta\nu = \nu_{n+1} - \nu_n$. Then when n is greater than 4 or 5 we have approximately

$$\Delta\nu = \frac{Ze}{4\pi a^{3/2}A^{1/2}m^{1/2}},$$

so that

$$a = \left(\frac{Ze}{4\pi\Delta\nu}\right)^{2/3}\frac{1}{(Am)^{1/3}}.$$

The nearly equally spaced energy levels found in many nuclei have spacings of 0·387 *MEV*. This energy difference corresponds to a frequency difference $\Delta\nu = 9{\cdot}382 \times 10^{19}$.

Putting in the numerical values of e, m and $\Delta\nu$, we get

$$a = 4{\cdot}65 \times 10^{-13}\left(\frac{Z^2}{A}\right)^{1/3}.$$

The following table gives the values of the nuclear radii given by this formula for several elements.

Element	A	Z	$a \times 10^{-13}$	$2{\cdot}47A^{1/3}$
Uranium	238	92	15 3	15·3
Gold	197	79	14·7	14·4
Silver	108	47	12·7	11·8
Cobalt	59	27	10 4	9 6
Aluminium	27	13	8 5	7·4
Beryllium	9	4	5·6	5·1

The last column gives the values of $2{\cdot}47A^{1/3}$.

These values of the nuclear radii given by the spherical-shell model agree as well as could be expected with the values found by other methods. On the spherical-shell model the nuclear particles are supposed to be arranged on the surface of a sphere, each one connected to three others near it by bonds analogous to the chemical valency bonds in molecules. Thus eight particles would be at the corners of a cube and twenty particles at the corners of a regular dodecahedron.

If s denotes the average distance between adjacent particles then $As^2 = 4\pi a^2$ approximately. Since a varies nearly as $A^{1/3}$, so that $a = a_1A^{1/3}$, we get $as^2 = 4\pi a_1{}^3$. The distance between the particles is therefore inversely as the square root of the nuclear radius. For uranium $s = 3{\cdot}5 \times 10^{-13}$ cm. and for beryllium $s = 6{\cdot}7 \times 10^{-13}$ cm.

16. Nuclear Fission.

When uranium is bombarded by neutrons it becomes radioactive, emitting β-rays. It was supposed that $_{92}U^{238}$ becomes $_{92}U^{239}$ and then emits a β-ray becoming $_{93}A^{239}$. If $_{93}A^{239}$ emitted another β-ray it would become $_{94}A^{239}$. $_{93}A^{239}$ and $_{94}A^{239}$ were called transuranic elements.

In 1939 Hahn and Strassmann found traces of radioactive barium in uranium which had been exposed to neutrons. The uranium was dissolved in nitric acid, a small quantity of barium chloride added to the solution of uranium nitrate, and then the barium precipitated as barium sulphate. It was found that part of the β-radioactivity of the uranium was precipitated along with the barium. A thorough investigation showed that the uranium contained radioactive barium and

also many other radioactive isotopes with mass numbers between about 72 and 158.

The radioactive isotopes found in the uranium in greatest quantities have mass numbers between about 90 and 102, and between 132 and 144

It appears that uranium atoms combine with a neutron and then break up into two parts, one part usually about 40 per cent heavier than the other. This process is called fission. The elements produced by fission include lanthanum, barium, caesium, xenon, iodine, tellurium and antimony, with atomic numbers from 57 to 51, and bromine, krypton, rubidium, strontium, ytterbium, zirconium, niobium and molybdenum with atomic numbers from 35 to 42.

Uranium has about 1·6 neutrons per proton but the stable isotopes of the elements formed by fission have only about 1·3 neutrons per proton. The isotopes formed by fission therefore have too many neutrons per proton for stability. The excess neutrons change into protons, with the emission of β-rays. This goes on until there are only about 1·4 neutrons per proton, and the nucleus is then stable.

For example, the following transitions occur when a barium atom ${}_{56}Ba^{144}$ is formed by fission:

$$ {}_{56}Ba^{144} \rightarrow {}_{57}La^{144} \rightarrow {}_{58}Ce^{144} \rightarrow {}_{59}Pr^{144} \rightarrow {}_{60}Nd^{144}. $$

${}_{60}Nd^{144}$ has 1·4 neutrons per proton and is stable.

Natural uranium is 99·3 per cent U^{238} and 0·7 per cent U^{235} with a trace of U^{234}. It is found that U^{235} fissions with neutrons of any kinetic energy from zero up, but U^{238} only fissions with neutrons with 1·5 *MEV* or more. Thorium and other elements with atomic numbers equal to or greater than 90 also fission with neutrons of sufficient energy.

A small number of neutrons is emitted when an atom fissions. Uranium atoms give between two and three. A very large amount of energy is released by fission. It appears mainly as kinetic energy of the particles formed. This energy can be calculated from the loss of mass. The atomic weight of U^{235} is 235·12 with $O^{16} = 16$. The elements formed by fission have atomic weights less than integers by about 0·05 and the two neutrons have atomic weights 1·00893. The loss of mass is therefore $0{\cdot}12 + 2 \times 0{\cdot}05 - 2 \times 0{\cdot}00893$, which is 0·20 mass units. The energy is therefore $0{\cdot}20 \times 931 = 188$ *MEV*. The energy has been found experimentally by measuring the heat produced when a known number of U^{235} atoms fission. It was found to be 185 *MEV* per uranium atom.

The energy released by the most energetic chemical reactions is never more than 10 electron volts per atom, so fission gives about 20 million times as much energy as any chemical reaction.

If a neutron goes into a large mass of pure uranium 238 with kinetic energy greater than 1·5 *MEV* it will collide with a uranium nucleus and cause it to fission. The nucleus will emit two neutrons which will cause

two more nuclei to fission. Four more neutrons will be emitted causing four more fissions. The four fissions will produce eight fissions and so on. This is called a chain reaction The chain reaction will proceed with great velocity, so that, in a very small fraction of a second, a large part of the uranium will fission with an enormous release of energy. The mass of uranium will explode Such a chain reaction is used in atomic bombs.

It is found that a chain reaction requires a mass of fissionable material greater than a certain value called the critical mass. Neutrons released inside fissionable material travel, on the average, an appreciable distance before they enter a nucleus and cause it to fission. With a small piece of material many of the neutrons escape outside the material and so do not produce fission The piece of material must be big enough so that one neutron released in it, on the average, produces more than one neutron by fission.

The critical mass for uranium 238 is very large, probably many tons, so it is not suitable for making atomic bombs The critical mass for uranium 235 is much smaller, only a few pounds, so it is suitable for making atomic bombs. Natural uranium contains about 0·7 per cent of uranium 235. The chemical properties of U^{235} are identical with those of U^{238}, so they cannot be separated by chemical processes. Uranium hexafluoride is a volatile liquid and $U^{235}F_6$ is slightly more volatile than $U^{238}F_6$. The $U^{235}F_6$ can therefore be partially separated from $U^{238}F_6$ by repeated fractional distillation. In this way uranium containing 10 to 15 per cent of U^{235} can be obtained. Pure U^{235} can then be obtained by separating the U^{235} from the U^{238} with mass spectrographs. In this way enough pure U^{235} has been obtained to make atomic bombs.

U^{238} does not fission with neutrons having kinetic energy less than 1·5 *MEV*, but it combines with slow neutrons forming ${}_{92}U^{239}$. This isotope is unstable and emits a β-ray forming ${}_{93}Ne^{239}$, a new element called neptunium. Neptunium is also unstable and emits a β-ray forming ${}_{94}Pl^{239}$, another new element called plutonium. Plutonium is a radioactive element emitting alpha rays. It is found that plutonium has a small critical mass like U^{235} and so is suitable for making atomic bombs. Plutonium has chemical properties quite different from those of uranium and so can be easily separated from it by chemical precipitation.

Plutonium is manufactured on a large scale with apparatus called piles. A pile is a large mass of graphite in which pieces of uranium are embedded. The graphite serves to slow down the fast neutrons emitted by the fission of the U^{235} in the uranium.

The neutrons emitted by the fission of the U^{235} diffuse through the graphite, colliding with the carbon nuclei. At each collision a neutron gives up some of its kinetic energy so that after about 100 collisions its energy is very small. The slow neutrons diffuse back into the uranium and about half of them enter U^{235} nuclei, causing fission, while about

half of them enter U^{238} nuclei, forming plutonium atoms. The pile must be large enough so that the number of neutrons which escape from the pile is only a small fraction of those produced by the fission of the U^{235}. If the fraction of the neutrons which escape is not too large, the number of neutrons in the pile will increase exponentially with the time, so that the rate of generation of heat by fission will also increase exponentially. The activity of the pile can be controlled by means of rods of cadmium or boron steel placed in holes in the pile. The rods absorb neutrons strongly. If the rods are pushed into the holes beyond a certain point the activity of the pile falls exponentially to zero. If the rods are pulled out the activity increases exponentially and by adjusting the rods the activity can be kept constant at any desired value.

The heat generated in the pile is removed by coils of aluminium pipe through which cold water is circulated. Such piles have been operated so as to generate one million kilowatts of power in the form of heat energy.

After the pieces of uranium have been in the pile for some time they are taken out and the plutonium in them extracted. The uranium is dissolved in acid, and the plutonium precipitated from the solution. The uranium from the pile contains all the products of the fission of the U^{235}, most of which are strongly radioactive emitting β- and γ-rays. It is, of course, necessary to protect the operators of the pile from the neutrons, β- and γ-rays emitted.

Atomic bombs have been made with uranium 235 and with plutonium. If two pieces of plutonium, each of say $\frac{3}{4}$ of the critical mass, are very quickly brought into contact the resulting mass is 50 per cent greater than the critical mass and so explodes. The explosion stops when the unexploded material is equal to the critical mass. The energy released is largely in the form of heat radiation. It is estimated that one atomic bomb is equivalent to about 25,000 tons of T.N.T.

17. Meson Theory of Nuclear Forces.

According to the photon theory of light the energy of the electromagnetic field is the energy of photons. The energy of a photon is equal to $h\nu$, where ν is the frequency. In a static field ν is zero so there must be an infinite number of photons of infinitesimal energy. It is supposed from this point of view, that electrons emit photons which are immediately absorbed by other electrons. The repulsion between two electrons may be regarded as an exchange force due to each electron emitting photons which are absorbed by the other one. The scalar potential ϕ in empty space satisfies the wave equation

$$\Delta\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = 0.$$

In electrostatics ϕ is constant and $\Delta\phi = 0$. The solution of this

equation is $\phi = \sum_n \frac{e_n}{r_n}$ where e_n is the charge at a distance r_n from the point at which ϕ is to be calculated. This makes the force between two electrons equal to e^2/r^2 and their potential energy e^2/r.

The force between two nuclear particles at very small distances is much greater than the force between two electrons, but it falls off very rapidly as the distance increases. Yukawa in 1935 proposed a theory of the forces between nuclear particles analogous to the above exchange-force theory for electrons.

Yukawa supposes that a proton in a nucleus emits a particle called a meson with a charge $+e$ and rest mass μ now believed to be about 200 times that of an electron. The meson is immediately absorbed by a neutron. Thus the proton changes into a neutron and the neutron into a proton. The protons are continually changing into neutrons and the neutrons into protons. Also a neutron may emit a meson with charge $-e$ so becoming a proton, and the negative meson may combine with a proton changing it into a neutron.

It is also supposed that neutrons and protons may emit neutral mesons. Thus the force between two neutrons or two protons may be due to each particle emitting neutral mesons, which are absorbed by the other particle, in the same way that the force between two electrons is due to each electron emitting photons which are absorbed by the other electron.

On this theory a proton may be regarded as a proton surrounded by a neutral meson cloud, or as a neutron with a positive meson cloud, or as a mixture of these two states.

In the same way a neutron may be regarded as a neutron with a neutral meson cloud or as a proton with a negative meson cloud or as a mixture of these two.

The emission of β-rays by radioactive atoms is supposed to be due to negative mesons disintegrating into electrons and neutrinos.

The wave equation for mesons, analogous to $\Delta\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = 0$ for photons, may be obtained as follows:

The momentum of a particle of rest mass μ and mass m is given by the equation $m^2v^2 = c^2(m^2 - \mu^2)$, obtained in Chap. I. The energy $E = mc^2 = -\frac{h}{2\pi i}\frac{\partial}{\partial t}$, so that $c^2m^2 = \frac{E^2}{c^2} = \frac{1}{c^2}\left(-\frac{h}{2\pi i}\right)^2\frac{\partial^2}{\partial t^2}$ and $m^2v^2 = p_x^{\,2} + p_y^{\,2} + p_z^{\,2}$, so that $m^2v^2 = \left(\frac{h}{2\pi i}\right)^2\Delta$. The equation for m^2v^2 therefore gives $\Delta\phi - \left(\frac{\mu c}{\hbar}\right)^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = 0$, where $\hbar = \frac{h}{2\pi}$. With $\frac{\partial^2\phi}{\partial t^2} = 0$ this becomes $\Delta\phi = \left(\frac{\mu c}{\hbar}\right)^2\phi$, analogous to $\Delta\phi = 0$ for photons.

The solution of this equation is $\phi = \frac{A}{r} e^{-\frac{\mu c}{h} r}$ where A is a constant. With $\mu = 200\ m_0$, where $m_0 = 9{\cdot}1 \times 10^{-28}$ the rest mass of an electron, we get

$$\frac{\mu c}{h} = \frac{200 \times 9{\cdot}1 \times 10^{-28} \times 3 \times 10^{10} \times 2\pi}{6{\cdot}62 \times 10^{-27}} = 5{\cdot}2 \times 10^{12}$$

The following table gives values of $(\phi/A)10^{-14}$ for several values of r.

r	$(\phi/A)10^{-14}$
10^{-16}	100
10^{-15}	10
10^{-14}	0·95
10^{-13}	0·06
10^{-12}	0·000073

Yukawa's theory is made rather more plausible by the discovery of particles with masses about 200 times that of an electron in cosmic rays. These particles are discussed under cosmic rays (p. 181).

Yukawa's theory makes the potential energy of the attraction between two nuclear particles nearly inversely as the distance between them when the distance is less than 10^{-14} cm., and very small for distances greater than 10^{-12} cm. The average distance between the particles in a nucleus is about 4×10^{-13} cm. With $r = 4 \times 10^{-13}$ cm $(\phi/A)10^{-14}$ is equal to 1/324, so that $\phi = \frac{A}{324} \times 10^{14}$. Now ϕ per particle in a nucleus is about 10 *MEV* or $1{\cdot}6 \times 10^{-5}$ ergs. This makes $A = 5{\cdot}2 \times 10^{-17}$ so that $\sqrt{A} = 7{\cdot}2 \times 10^{-10}$. The attraction between two nuclear particles at very small distances is therefore about the same as that between two electric charges of opposite sign, one $+15e$ and the other $-15e$, since $e = 4{\cdot}8 \times 10^{-10}$.

REFERENCES

1. *Elements of Nuclear Physics*. Rasetti.
2. *Radioactivity and Nuclear Physics*. Cork.
3. "Nuclear Physics A." Bethe and Bache. *Reviews of Modern Physics*, April, 1936.
4. "Nuclear Physics B." Bethe. *Reviews of Modern Physics*, April, 1937.
5. "Nuclear Physics C." Livingston and Bethe. *Reviews of Modern Physics*, July, 1937.
6. "Nuclei formed by Fission." *Reviews of Modern Physics*, October, 1946.
7. *Meson theory of nuclear forces*. W. Pauli, 1946.
8. "Recent research in meson theory." G. Wentzel. *Reviews of Modern Physics*, January, 1947.

CHAPTER XIII

Gaseous Ions

1. Mobility of Ions.

When gases are made to conduct electricity by means of X-rays, ultra-violet light, or other agencies, it is supposed that the conductivity is due to the presence of minute electrically charged particles which are called ions. The properties of these ions is the subject to be discussed in this chapter.

The velocity of ions due to an electric field depends on the temperature, pressure, and nature of the gas in which they are moving. It is proportional to the strength of the electric field, and the velocity due to a field of unit strength, usually one volt per centimetre, is called the mobility of the ions. A group of a large number of ions in a uniform electric field moves with a definite velocity, that of the centre of mass of the group, but this is not true of each ion in the group. The ions diffuse through the gas when there is no field, and in a field they still diffuse, so that the displacements due to diffusion are added to those due to the motion caused by the electric field. In a strong field the motion due to diffusion may be very small compared with that due to the field The mobility of a positive ion will be denoted by k_1 and that of a negative ion by k_2.

If a gas contains n_1 positive ions per unit volume, and n_2 negative ions, the current density in it due to an electric field of strength F is

$$i = (n_1 e_1 k_1 + n_2 e_2 k_2)F,$$

where e_1 is the charge on a positive ion and e_2 that on a negative ion. The products $e_1 k_1$ and $e_2 k_2$ are both positive, because changing the sign of e also changes the sign of k. Practically all gaseous ions are found to carry charges of the same magnitude of either positive or negative electricity. This ionic charge is equal to the charge on one hydrogen ion in a solution, or to the charge carried by one electron. It will be denoted by e. The current density is therefore given by

$$i = (k_1 n_1 + k_2 n_2)eF,$$

where both k_1 and k_2, as well as e, are now regarded as positive.

2. Measurement of Ionic Mobilities. Zeleny's Method.

Many determinations of the ionic mobilities k_1 and k_2 have been made in various gases at different pressures and temperatures. It is only possible to describe a few of them here.

In 1900 Zeleny made some accurate measurements of k_1 and k_2 for the ions produced by X-rays in several gases at atmospheric pressure.

The gas was passed at an uniform rate through a tube TT′ of circular cross-section, along the axis of which was a long insulated cylindrical electrode CF (fig. 1). This electrode was divided into two parts by a narrow gap at DE. A narrow beam of X-rays SAB was passed across the tube from a source S so that positive and negative ions were produced in the gas over the cross-section AB of the tube. The electrode CF was kept at zero potential and the tube TT′ charged to a potential V by means of a battery. As the ions are carried along the tube in the stream of gas, those of one sign move towards the tube and the others towards the electrode. The ions produced at the surface of the tube which move to the electrode strike the electrode farthest from AB. If V is small some ions will reach EF, and as V is increased the number reaching EF diminishes to zero. The value of V just big enough to prevent any ions from reaching EF was determined. With this value of V the ions starting from the surface of the tube at AB move across to E. The ions reaching EF were detected by means of a quadrant electrometer to which it was connected.

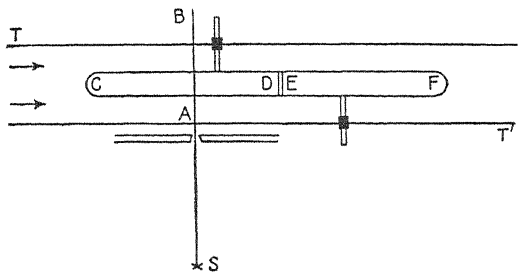


Fig 1

Let the radius of the tube be b and that of the electrode a, and let F be the strength of the electric field from the tube towards the electrode. Then F is inversely as r, so that $F = V/\left(r \log \frac{b}{a}\right)$, provided there are not enough ions present to sensibly disturb the field. If v is the velocity of the gas stream, then in a time dt an ion will be carried along with the stream a distance $dx = vdt$, where x is the distance along the tube from AB, and will move in towards the electrode a distance $dr = kFdt$. Hence $dr = kFdx/v$, or

$$2\pi vrdr = \frac{2\pi Vkdx}{\log(b/a)}.$$

Integrating along the path of the ions from $r=b$, $x=0$ to $r=a$, $x=d$, where d is the distance between EF and AB, we get

$$Q=\int_a^b 2\pi vr dr=\frac{2\pi Vk}{\log\frac{b}{a}}d,$$

where Q is the volume of gas flowing past AB in unit time. Hence

$$k=\frac{Q\log b/a}{2\pi Vd}.$$

3. **Langevin's Method.**

Another very good method of finding ionic mobilities was used by Langevin. The space between two parallel plates was filled with the gas, and ions were produced by a single flash of X-rays got by breaking the primary circuit of an induction coil connected to the X-ray tube. An electric field was maintained between the plates which was reversed in direction at a time interval t after the flash. The charge received by one of the plates was measured with a quadrant electrometer. This charge Q varies with the interval t, and the ionic mobilities can be found from the relation between Q and t. Let the distance between the plates be l and the strength of the electric field X. Suppose that the charge on the positive ions produced by the flash of X-rays is q per unit volume of the gas, and that the positive ions move towards the plate connected to the electrometer during the interval t. The charge received by the plate during the interval t will then be Aqk_1Xt, where A is the area of cross-section of the flash of X-rays. When the field is reversed the negative ions remaining between the plates will go to the plate, which will therefore receive a negative charge $Aq(l-k_2Xt)$, so that we have

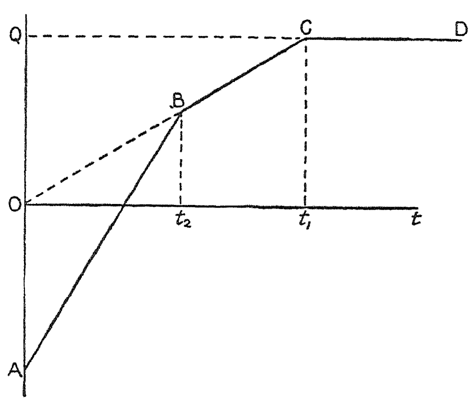


FIG. 2

$$Q=Aqk_1Xt-Aq(l-k_2Xt)=Aq\{Xt(k_1+k_2)-l\}.$$

This expression is correct so long as k_1Xt and k_2Xt are less than l. If k_2 is greater than k_1, as is usually the case, then when t is greater than $t_2=l/k_2X$ we shall have

$$Q=Aqk_1Xt,$$

and when t is greater than $t_1 = l/k_1X$ we shall have

$$Q = Aql$$

The relation between Q and t is shown in fig. 2. When $t = 0$, $Q = -Aql$, since all the negative ions go to the plate, and when $t > t_1$ then $Q = Aql$, since all the positive ions go to the plate. By measuring Q for a series of values of t and plotting the results the values of t_1 and t_2 can be found. k_1 and k_2 are then given by $k_1 = l/Xt_1$ and $k_2 = l/Xt_2$. We have assumed that the charges in the gas are not sufficient to appreciably modify the electric field and that the time t is so short that no appreciable recombination of the ions takes place. Actually, some recombination does occur, which alters the relation between Q and t but does not affect the times t_1 and t_2 corresponding to the breaks in the curve at C and B.

4. **Rutherford's Method.**

Another excellent method of measuring ionic mobilities was first used by Rutherford.

Two parallel plates at a distance l apart are used with the gas between them. Ions, of one sign only, are produced at the surface of one of the plates. If now an electric field X is maintained between the plates for a time t and then reversed, no ions will reach the other plate unless t is greater than l/kX. Thus k can be determined by finding the value of t at which the other plate begins to receive a charge. This can be done by varying either t or l. The electric field may be applied for the time t a great many times, the field being reversed, after each application, for long enough to return all ions in the gas to the plate from which they started. In this way the charge to be detected can be greatly increased. The field is applied and reversed at regular intervals by a rotating commutator.

An alternating electric field may be used in this method. Thus if $X = X_0 \sin(2\pi t/T)$ the distance the ions move is given by

$$\int_0^{T/2} k\,X dt = kX_0T/\pi.$$

Hence if $kX_0T/\pi = l$ the plate will just begin to receive charge. By varying l it is easy to find the value of kX_0T/π.

An error may arise owing to diffusion of the ions. An ion in a gas when there is no electric field does not remain at rest but moves about in an irregular manner first in one direction and then in another. Thus some ions will move a greater distance than kX_0T/π and some not so far. Those which move farther may not get back to the plate when the field is reversed and so may get still farther from it the next time they move away, and so on. The method therefore tends to give too high a value of the mobility. Zeleny's and Langevin's methods may also be affected by a similar error due to diffusion. This error may

be reduced by superposing on the alternating field a steady field $\overline{X}$ in the direction tending to retard the motion of the ions from the plate at which they start. In this case

$$X = X_0 \sin \frac{2\pi t}{T} - \overline{X},$$

so that the distance the ions move across is

$$k\int_{t_1}^{t_2}\left(X_0 \sin \frac{2\pi t}{T} - \overline{X}\right) dt.$$

Here t_1 and t_2 are the times between which X is positive. This is approximately

$$kT\left(\frac{X_0}{\pi} - \frac{\overline{X}}{2} + \frac{\overline{X}^2}{2\pi X_0}\right),$$

when $\overline{X}$ is not a large fraction of X_0.

The supply of ions at the surface of one of the plates may be produced by allowing ultra-violet light to fall on the plate. This causes the emission of negative ions by the plate. Another way is to have a hole in the plate covered with fine wire gauze. The gas behind the gauze is ionized by X-rays, and some of the ions of one sign are made to go through the gauze by means of a weak electric field.

5. Results of Various Experimenters.

The following table contains some of the results on ionic mobilities which have been obtained. k_1 and k_2 are expressed in centimetres per second for 1 volt per centimetre

Gas	Pressure in Mm. of Mercury	k_1	k_2	Method.	Observer
Air	760	1 36	1 87	Zeleny's	Zeleny
Hydrogen ..	760	6 70	7·95	,,	,,
Air	75	14 8	21·90	Langevin's	Langevin
,, ...	200	5 45	7·35	,,	,,
,, ...	415	2 61	3 31	,,	,,
,, .	760	1 40	1·70	,,	,,
,, . ..	1430	0 75	0 90	,,	,,
Carbon monoxide	760	1·10	1·14	,,	Wellisch
Carbon dioxide .	760	0 81	0·85	,,	,,
Nitrous oxide .	760	0 82	0·90	,,	,,
Sulphur dioxide	760	0·44	0 41	,,	,,
Methyl iodide .	760	0 21	0 22	,,	,,
Ethyl iodide ..	760	0 17	0 16	,,	,,
Helium	760	5·09	6 31	Rutherford's	Franck & Pohl
Argon .. .	760	1 37	1 70	,,	Franck
,, (pure) .	760	1 37	206·0	,,	,,
Nitrogen ..	760	1 27	1·84	,,	,,
,, (pure) ..	760	1·27	144·0	,,	,,

It appears that the mobility of the negative ions is generally rather greater than that of the positive ions. The velocity of the positive ions is nearly inversely as the gas pressure. In very pure argon and nitrogen Franck found that the negative ions had much larger velocities than in these gases not specially purified. A small quantity of oxygen added to the pure argon reduced k_2 to a much smaller value.

Zeleny found that water vapour diminishes the velocity of the negative ions in air and hydrogen, making it nearly equal to that of the positive ions. The velocities of the negative ions in air, CO_2, and hydrogen at pressures down to a few millimetres have been determined by Lattey and Tizard, using a modification of Rutherford's method. They found that the velocity is approximately a function of X/p or the ratio of the electric field strength to the pressure. In the case of the positive ions the velocity v_1 is given by $v_1 = k_1X/p$, where k_1 is the mobility at unit pressure. For the negative ions $v_2 = k_2X/p$, when X/p is small, but as X/p increases the velocity v_2 increases more rapidly than X/p.

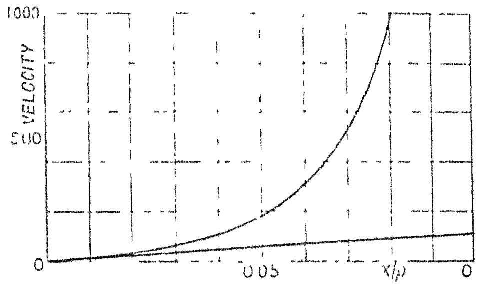


Fig. 3.

The relations between v_1 and v_2 and X/p in dry air are shown in fig. 3. When X is expressed in volts per centimetre and p in millimetres of mercury, $v_1 = 1080X/p$, and when X/p is less than about 0 01, $v_2 = 1350X/p$, but when X/p is equal to 0 08 the velocity v_2 is equal to 1000 cm./sec., which is ten times $1350X/p$.

6. Theory of Ionic Velocities.

It will be convenient now to consider the theory of the motion of ions through a gas in an electric field. The ions are supposed to move about in the gas with a high velocity and to collide with the gas molecules just as the uncharged molecules do. The average kinetic energy of an ion in the absence of an electric field will be equal to the average kinetic energy of a gas molecule, so that we have $\frac{1}{2}mV^2 = \frac{1}{2}m'V'^2$, where m and m' are the masses of an ion and a molecule, and V^2 and V'^2 the average values of the squares of their velocities.

In an electric field the ions acquire an average velocity of drift u in the direction of the field. The average momentum of an ion is then mu, while when there is no field their average momentum is zero, since there are then as many moving in any direction as in the opposite direction. If e is the charge on an ion then in a field of strength X the force on it is Xe, so that the field gives the ion momentum at the rate Xe per unit time. The ions lose the momentum they receive from the field by colliding with the molecules, and in a steady field the rate at which they lose momentum is equal to that at which they receive it. If λ denotes the mean free path of an ion, or the average distance between its collisions with the gas molecules, then the number of collisions it makes in unit time is approximately V/λ. This is not exact, because V is not the average velocity but the square root of the mean square of the velocity. If on the average an ion loses a fraction f of its average momentum mu at each collision, the momentum it loses in unit time is $Vfmu/\lambda$. Hence in a steady state

$$Xe = Vfmu/\lambda,$$

so that

$$u = \frac{Xe\lambda}{fmV}.$$

In the case of an ion having a mass very small compared with that of a gas molecule, f will be nearly unity, so that

$$u = \frac{Xe\lambda}{mV},$$

but in the case of a heavy ion having a mass larger than that of a gas molecule f will be a small fraction.

7 Mobility and Coefficient of Diffusion.

The mobility of an ion may be obtained from its coefficient of diffusion K. We have

$$-K\frac{dp}{dx} = pv,$$

where p denotes the partial pressure of the ions, and v the velocity of diffusion in the x direction. Also if there are n ions per unit volume dp/dx may be regarded as the force on the ions in unit volume which drives them along with the velocity v. In an electric field X the force on the n ions is Xen, and this gives them a velocity kX. Hence, assuming the velocity proportional to the driving force, we have

$$\frac{v}{kX} = \frac{dp/dx}{Xen} = \frac{pv}{KXen},$$

or

$$k = \frac{Ken}{p}.$$

But $p = nR_1T$, where R_1 is the gas constant for one molecule, so that

$$k = \frac{Ke}{R_1T}.$$

If $\mathfrak{N}$ is the number of molecules in 1 gm. molecule or mol, then

$$k = \frac{K\mathfrak{N}e}{RT},$$

where R is the gas constant for 1 mol. We have $\mathfrak{N}e = 9650$ electromagnetic units, and $R = 8.32 \times 10^7$, so that at 300° K. we get, after multiplying by 10^8 to convert electromagnetic units of field strength into volts per centimetre, $k = 39K$. The coefficient of diffusion of oxygen molecules diffusing in oxygen at 760 mm. and 300° K is approximately 0·19, which gives $k = 7·4$. For hydrogen diffusing through hydrogen $K = 1·3$, so that $k = 51$. The values found for k_1 in oxygen and hydrogen are 1·4 and 6·7, and for k_2 1·9 and 8·0 cm. per second for 1 volt per centimetre. These values are from four to eight times smaller than the values just calculated. This is believed to indicate that the ions are not single molecules carrying a charge e but clusters of several molecules held together by the electrical field of the charge.

In the case of the negative ions the clusters evidently diminish in size as X/p increases, and when X/p is greater than about 0·1 the negative ions are reduced to electrons. The motion of electrons in gases is considered in another chapter.

8. Townsend's Determination of Coefficients of Diffusion.

The coefficients of diffusion of the ions produced in gases by X-rays or other sources of ionization were determined by J. S. Townsend. Townsend passed the gas containing ions through narrow metal tubes, and determined the fraction

of the ions lost by diffusion to the walls of the tubes. He also determined the ratio of the mobility to the coefficient of diffusion by the method described in the chapter on the motion of electrons in gases. This ratio, as we have seen above, is equal to Ne/RT. If e is equal to the charge on one univalent ion in a solution then we know that Ne is equal to 96,500 coulombs, so that Townsend's determinations of k/K enabled him to prove that the charges on gaseous ions are equal to the charge on a univalent ion in solutions. He found Ne for positive ions always nearly equal to 96,500 coulombs, and the same was true for negative ions when they had small mobilities, showing that they were clusters of molecules like the positive ions. The values of the coefficients of diffusion of ions found by Townsend by the narrow tube method are as follows for ions produced by X-rays:

Gas.	Positive Ions.	Negative Ions.
Air	0·028	0·043
Oxygen	0·025	0·040
Carbonic acid	0·023	0·026
Hydrogen	0·123	0·190

Townsend also measured the coefficient of diffusion of the ions produced in air by several different sources of ionization and found the following values:

	Positive Ions.	Negative Ions.
X-rays	0·028	0·043
Radium rays	0·032	0·043
Ultra-violet light	—	0·043
Point discharge	0·024	0·035

He found that the coefficients of diffusion were inversely as the gas pressure from 200 mm. to 772 mm. of mercury.

Salles has made measurements of the coefficients of diffusion of ions in several gases by Townsend's tube method, and obtained results agreeing approximately with those found by Townsend.

9. Another Method of Determining the Charge on One Mol of Gaseous Ions.

The charge carried by 1 gm.-molecule or mol of gaseous ions was determined by the writer by an entirely different method. Air containing a small amount of a solution of an alkali salt in suspension in the form of fine spray was passed through a long platinum tube heated in a furnace to a high temperature. The salt is volatilized in the tube and is ionized at the high temperature. The charge carried by the ions formed was found by measuring the current between the platinum tube and a cylindrical electrode placed along the axis of the tube. It was found that above about 1300° C. and with a potential difference between the electrode and tube of more than 600 volts the current obtained was nearly independent of the temperature and potential difference. This saturation current was measured for a number of different alkali salts, and was found to be proportional to the amount of salt passing into the tube in unit time and inversely

proportional to the electrochemical equivalent of the salt. The quantity of electricity carried by the ions formed from 1 gm. equivalent of any of the salts used was found to be approximately 98,600 coulombs, which agrees with the value 96,500 found in solutions within the limits of error. The salts used were CsCl, Cs_2CO_3, RbI, Rb_2CO_3, RbCl, KCl, KI, KBr, KF, K_2CO_3, NaI, NaCl, NaBr, Na_2CO_3, LiI, LiCl, LiBr, Li_2CO_3.

The measurements which have been made on the ratio of the charge to the mass of positive rays, which are discussed in the chapter on positive rays, show that the positive ions formed in gases at low pressures may have charges which are small multiples of the ionic charge e, but in gases at atmospheric pressure at any temperature gaseous ions are found to have charges equal to e. Probably in electric discharges with very intense fields ions having charges which are small multiples of e may be formed even at high pressures.

The fact that ions formed under such varied conditions in liquid electrolytes of any kind, in gases by any source of ionization at any pressure and at any temperature and from any kind of compound or element, always have charges which are small but exact multiples of the ionic charge e, shows conclusively that electricity has an atomic constitution. That is to say that electricity is composed of small exactly equal atoms of electricity and that the charge on the negative atoms is equal to the charge on the positive atoms.

10. Recombination of Ions.

When positive and negative ions are present in a gas they recombine and form electrically neutral molecules. The rate of recombination is proportional to the number n_1 of positive ions and to the number n_2 of negative present in unit volume. Thus we have

$$-\frac{dn_1}{dt} = -\frac{dn_2}{dt} = \alpha n_1 n_2,$$

where α is a constant called the coefficient of recombination. If $n_1 = n_2 = n$, then

$$-\frac{dn}{dt} = \alpha n^2,$$

which gives $\frac{1}{n} - \frac{1}{n_0} = \alpha t$, where n_0 is the value of n when $t = 0$. We are supposing here that no ions are removed from the gas except by recombination.

The coefficient of recombination of ions in gases has been determined by Rutherford, Townsend, Langevin, and others by measuring the rate at which the ions disappear. The values of α/e found in different gases at atmospheric pressure vary between 3000 in hydrogen and 3400 in air, with e in electrostatic units. The factor e^{-1} appears because the quantity actually measured is not the number of ions n but the charge $q = ne$ which they carry. Thus we have

$$\frac{e}{q} - \frac{e}{q_0} = \alpha t,$$

so that
$$\frac{\alpha}{e} = \frac{1}{t}\left(\frac{1}{q} - \frac{1}{q_0}\right).$$

Langevin found that α/e is nearly proportional to the pressure of the gas between 760 and 150 mm. of mercury. At pressures of several atmospheres in air and CO_2 he found α/e to be nearly inversely as the pressure. At very low pressures the coefficient of recombination probably becomes very small, but no determinations of it have been made at low pressures.

A simple theory of recombination has been given by Langevin which is approximately correct in gases at high pressures. Consider a sphere of radius r in a gas with a positive ion at its centre. If there are n_2 negative ions per unit volume, the number of negative ions entering the sphere per unit time owing to the electric field of the positive ion will be $4\pi r^2 n_2(k_1 + k_2)\frac{e}{r^2}$, since a positive and a negative ion move towards each other with the relative velocity $(k_1 + k_2)\frac{e}{r^2}$, and when r is small there will very seldom be more than one negative ion at the surface of the sphere. If there are n_1 positive ions present in unit volume the total number of negative ions moving up to positive ions per unit volume in unit time will therefore be $4\pi e n_1 n_2(k_1 + k_2)$, so that if we suppose that recombination occurs in every case we get

$$\alpha = 4\pi e(k_1 + k_2)$$

At high pressure this formula gives values of α/e in agreement with those observed. For example, in carbonic acid Langevin found that at three atmospheres pressure

$$\frac{\alpha}{4\pi e(k_1 + k_2)} = 0{\cdot}97.$$

At pressures below one atmosphere α is much smaller than $4\pi e(k_1 + k_2)$. This is due to the fact that an ion may move up to another one and move round it and then diffuse away without recombining. Unless the ion loses kinetic energy by collisions when close to the other ion it is unlikely to recombine. The electric field e/r^2 of an ion is too weak, except at very small distances, to appreciably affect the diffusion of another ion.

11. Formation of Clouds on Ions.

Clouds consisting of minute drops of water are formed in gases by the condensation of supersaturated water vapour on dust particles or other nuclei. In the absence of such nuclei no cloud is formed unless the degree of supersaturation is very high. C. T. R. Wilson discovered that the ions formed in gases by X-rays or other sources of ionization can act as nuclei for the condensation of supersaturated water vapour. The apparatus used by C. T. R. Wilson is shown in fig. 4. BB is a glass tube about 20 cm. long and 4 cm. in diameter which is closed at its lower end by a rubber stopper R. Another glass tube CC with its upper end closed slides freely up and down inside BB. A tube DD passing through the stopper connects the inside of CC through a valve S, opened by pulling T, to a large bottle E, which is kept exhausted by means of a tube P leading to a pump. The tube BB is about half filled with water as shown.

The upper end of BB is connected to a mercury manometer F through a stopcock, and air can be withdrawn or admitted through a tube G packed with glass wool to remove dust particles. A bulb A is connected to the top of BB by means of a wide tube. If the valve S is shut and the stopcock Q opened, the sliding tube CC, which acts as a piston, may be raised to any desired point by removing some air through G. If Q is then shut and S opened a vacuum is produced under

the piston, which very suddenly moves down until it strikes the rubber stopper In this way the air in A is made to expand very rapidly. The change of pressure due to the sudden expansion can be measured with the manometer F. The expansion produces supersaturation of the water vapour in A, and if any dust particles are present a cloud is formed which falls slowly and finally settles on the glass walls so removing the dust particles from the gas.

By repeating this process two or three times all dust can be got rid of. In dust-free air no cloud is formed when the expansion ratio is less than 1·25. With ratios between 1·25 and 1·38 a small number of drops, about 100 per cubic centimetre, appear, and with expansion ratios above 1·38 a very dense cloud is formed which is evidently due to condensation on the molecules.

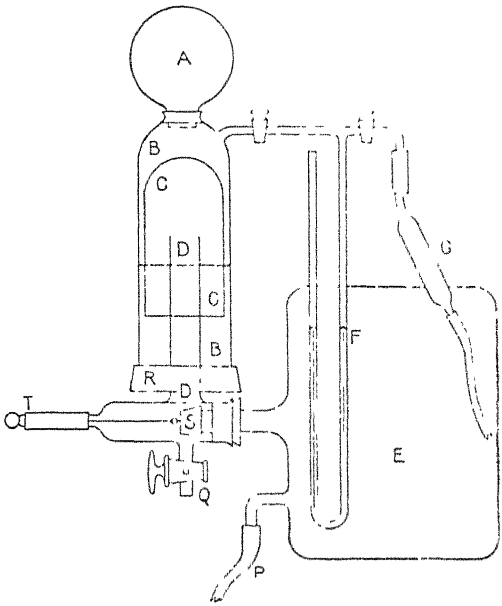


Fig 4

If X-rays are passed through the bulb A then a cloud is formed with expansion ratios between 1·25 and 1·38. This effect persists for a short time after the rays are cut off If an electrode is sealed into A, it is found that charging it to a few hundred volts stops the formation of any cloud after the rays are cut off. This shows that the cloud produced by the rays is due to condensation on the ions formed by the rays. C. T. R. Wilson found that the clouds are formed more easily on negative than on positive ions. An expansion of 1·25 gives a cloud with negative ions, but 1·31 is required with positive ions.

12. Photographs of Tracks of α- and β-rays.

C. T. R. Wilson found that it is possible to make visible and to photograph the tracks of individual α-rays and β-rays through gases by forming a cloud on the ions produced by the rays. For this purpose a special form of the expansion apparatus is used. The expansion chamber is a tube about 15 cm. in diameter closed at the top with a glass plate. The top of the piston is also a flat plate, painted black

and nearly as large as the top of the expansion chamber. The distance between the top of the expansion chamber and the top of the piston is about 3 cm., and is suddenly increased when the piston is connected to the vacuum bottle as with the first apparatus. A rather large potential difference is maintained between the top and bottom of the expansion chamber to remove ions formed immediately. If a very minute particle of radium is supported on a rod inside the expansion chamber, then when the sudden expansion occurs the tracks of the α-rays appear as long narrow lines of cloud. To see them clearly it is necessary to illuminate the chamber strongly. The tracks of the α-rays are usually straight lines, but occasionally a track appears in which there is a sudden change of direction at one point which is usually near the end of the track. These tracks must be photographed immediately after they have been formed. They very quickly disappear. Two photographs are reproduced here of α-ray tracks through nitrogen, showing collision between an α-ray and a nitrogen atom. (See fig. 5, Plate facing p. 157.)

Harkins and also Blackett have taken photographs of enormous numbers of α-ray tracks Blackett obtained several photographs showing collisions between an α-ray and a nitrogen atom in which a hydrogen atom was knocked out of the nitrogen atom. The hydrogen atoms gave straight tracks, much narrower than the α-ray tracks and easily distinguishable from them. In such collisions the α-ray apparently remains inside the nitrogen atom and the nitrogen atom acquires sufficient velocity to produce a short track. The track of the α-ray therefore branches at the collision into a long straight very narrow track and a short thick track. Collisions are frequently observed in which an α-ray track branches into two similar short tracks. This is supposed to be due to the α-ray colliding with an atom and rebounding from it, so that the α-ray and the atom both produce a short track after the collision. By measuring the lengths of the branches the velocities of the colliding particles can be estimated, and it is found that in these collisions energy and momentum are conserved as for perfectly elastic bodies. When a narrow beam of X-rays is examined by this method two distinct types of track are observed. The electrons shot out of the atoms give tracks about 1 cm. or so long which are not straight but continually change in direction, especially towards the end of the track. These tracks start out sideways from the X-ray beam and more often with a forward velocity component than not. Besides these rather long tracks very short tracks are also observed which are believed to be due to the scattering of an X-ray quantum by an electron. The electron then gets only a small amount of energy and so only gives a short track. A short track is frequently associated with a long track which starts outside the X-ray beam and is supposed to be produced by the scattered quantum.

Thus C. T. R. Wilson's method enables effects due to single atoms and electrons to be observed and photographed, and has already led to most interesting and important results.

Direct Determination of the Ionic Charge

13. Up to about 1897 the value of the ionic charge e could only be obtained from the number of molecules in gases, as deduced from the viscosity and other properties of gases by the results of the kinetic theory. The product $\mathfrak{N}e$ where $\mathfrak{N}$ is the number of molecules in 1 gm.-molecule or mol of any gas, was known accurately from the measurements of the electrochemical equivalent of silver, but the possible error in the value of $\mathfrak{N}$ deduced from the kinetic theory of gases was 50 per cent or more. About that date there was nevertheless convincing evidence available proving the atomic nature of electricity. The facts of electrolysis alone showed that the average ionic charge in liquid electrolytes is always a small integral multiple of a definite unit of charge, and the fact that the rays in a beam of cathode rays are all deflected to the same extent by a magnetic field showed that the rays all have the same value of e/m and are not a mixture of different sorts of particles having constant properties only on the average. About that time it was shown by J. J. Thomson and others that electrons obtained from many different sources have identical properties, and it was clear that these electrons are constituents of the chemical atoms, which were therefore not indivisible particles. These conclusions were supported by an immense amount of new evidence during the next ten years, and the theory of the atomic nature of electricity was established beyond reasonable doubt before Millikan succeeded in measuring very small electric charges with sufficient accuracy to show that they were always exact multiples of an atomic unit. This result came therefore as a final confirmation of a well established theory, but was and is nevertheless of the highest interest and importance.

14. Townsend's Method.

The first experiments which could be regarded as a direct determination of the ionic charge were made by Townsend in 1897. The results he obtained were not very exact, but the method he devised contained several of the essential features of the methods used in subsequent investigations. Townsend showed that it was possible to make a direct determination of the ionic charge, and started the attack on the problem.

Townsend found that the gases evolved in the electrolysis of dilute sulphuric acid by rather large currents form dense clouds when bubbled through water. These clouds consist of minute drops of water containing a trace of some hygroscopic substance, probably H_2SO_4, and they are electrically charged. He measured the charge carried by

the cloud and the amount of water in it, and found the mass of the droplets by measuring the rate at which they fell through the gas. In this way he got the total number of drops in the cloud and the total charge on them. Dividing the charge by the number of drops he got the average charge per drop, which was about 3×10^{-10} electrostatic units. Later he found that all the drops were not charged with electricity of the same sign, and allowing for this got for the average charge per drop 5×10^{-10} e.s.u. Townsend supposed that the drops were formed on gaseous ions, and so regarded the result obtained as a determination of the ionic charge. The drops in Townsend's clouds all fell at nearly the same rate, since the upper surface of the cloud remained distinct as it fell. This showed that the majority of the drops were of equal size, and so justified to some extent the assumption that they all carried equal charges.

The size of the drops was obtained from the rate of fall by means of the theory, due to Stokes, of the motion of a sphere through a viscous fluid According to Stokes's theory the resistance to the motion of a sphere of radius a through a medium of viscosity μ, with small uniform velocity v, is equal to $6\pi\mu va$. It is assumed in deducing this expression that there is no slipping at the surface of the sphere and that $av\rho'/\mu$ is a small fraction, ρ' being the density of the medium For a sphere of water falling through air, a must be small compared with 0·006 cm. The weight of a sphere of density ρ is $\frac{4}{3}\pi a^3\rho g$, so that for a falling sphere we have

$$\tfrac{4}{3}\pi a^3 g(\rho - \rho') = 6\pi\mu va,$$

when the velocity v has attained a constant value. Hence

$$v = \frac{2}{9}\,\frac{\rho - \rho'}{\mu}\,ga^2,$$

which is the expression used by Townsend to get the radius a and so the mass $m = \frac{4}{3}\pi a^3\rho$ of his droplets.

The weak points in Townsend's determination of e are the assumption that all the drops carry equal charges and that these equal charges are equal to the ionic charge. It was not clear that only one ion went into each drop—in fact the way in which the charges got into the drops was not known.

15. Method of J. J. Thomson.

Soon after Townsend's experiments appeared, J. J. Thomson carried out a determination of e by a method which was based on essentially the same principles, but in which the droplets were formed by condensing water vapour on the ions formed in air by X-rays or radium rays.

It had been shown by C. T. R. Wilson that such ions act as nuclei for the condensation of water vapour, so that in J. J. Thomson's experiments it was clear that the charge measured was the ionic charge.

This celebrated investigation has therefore sometimes been regarded as the first direct determination of the ionic charge. J. J. Thomson obtained the cloud on the ions by means of C. T. R. Wilson's expansion apparatus described above, and got the mass of the drops from their rate of fall. He found the charge in the gas by measuring its electrical conductivity due to a small electric field. The conductivity is equal to $ne(k_1 + k_2)$, where n is the number of positive or negative ions per cubic centimetre, and k_1 and k_2 are the mobilities of the ions. The mobilities in air were known and so ne could be calculated. The total mass of water in the cloud was calculated from the expansion ratio, assuming the expansion to be adiabatic. In this way J. J. Thomson showed that e was nearly the same for ions produced by X-rays, radium rays, and ultra-violet light. The final result he got for e was $3{\cdot}4 \times 10^{-10}$ electrostatic units.

The weak points in these experiments are the assumptions that all the drops contain only one ion and that the total mass of water in the cloud is equal to that due to an adiabatic expansion. As a matter of fact the clouds soon evaporate as the air warms up, so that the mass of the drops is not constant.

16. Another Method of Determining the Ionic Charge.

A method of finding the charge on the drops which makes it unnecessary to determine the total mass of the cloud and the total number of drops was proposed by the writer in 1903. The velocity of a sphere moving through a viscous liquid is proportional to the force driving it, so that in a vertical electric field F which exerts a force on the drops equal to Fe the velocity of the drops will be changed. If v_1 is the rate of fall with $F = 0$, and v_2 the rate of fall in the field, then

$$\frac{v_1}{v_2} = \frac{mg}{mg + Fe},$$

where m is the mass of the drop, and the force Fe is reckoned positive when its direction is downwards. This gives

$$e = \frac{mg(v_2 - v_1)}{Fv_1}.$$

The mass m can be got from v_1 as in Townsend's and J. J. Thomson's experiments, so that e can be calculated from v_1 and v_2.

The apparatus used consisted of a C. T. R. Wilson expansion chamber containing two horizontal parallel electrodes between which a vertical electric field could be maintained by means of a battery giving up to 2000 volts. The electrodes were 3·5 cm. in diameter and about 5 mm. apart. The space between them was illuminated by a narrow parallel beam, and the cloud was observed on the axis of the electrodes where the field was practically uniform. The moist air in the expansion

chamber was ionized by X rays, which were usually cut off just before an expansion.

It was found that the clouds formed with no electric field consisted of drops which very nearly all fell at practically the same rate. The individual drops could be seen unless the cloud was very dense, and it was easy to see if there were any drops falling at different rates. This was not the case in the cloud near the axis where it was observed.

It was also found that the clouds formed in successive expansions under the same conditions all fell at practically the same rate, so that the plan adopted was to measure the time for the top of a cloud to fall from the upper electrode to the lower one, first without any field and then with a field. This gave better results than were obtained by measuring the rate of fall for part of the distance without any field and for the rest of the distance with a field applied. Owing to the evaporation of the drops the rate of fall is not uniform, so that it was better to measure the average rate over the whole distance without any field and then with a field. The fact that the clouds formed in successive expansions all fell at the same rate justified this procedure, which was adopted because it gave the best results and not because the method of measuring v_1 and v_2 on the same cloud was not tried.

It was found that in an electric field all the drops did not fall at the same rate. Several sets of drops could be seen and all the drops in each set fell at the same rate. Three such sets could usually be detected, and it was found that the charges on the drops in them were nearly in the ratio 1 : 2 : 3. This result showed clearly that the charges on the drops were multiples of an atomic unit, as they should be according to the atomic theory of electricity.

If the X rays were kept on during and after the expansion a few droplets were obtained carrying comparatively large charges. Some of these drops could be made to rise with a potential difference of only a few hundred volts. The charges on these drops were found to be rather large multiples of the ionic charge and they were not used to determine e, because owing to evaporation it did not seem to be possible to determine their charges with sufficient accuracy to establish the value of the integer expressing the charge in terms of e. It was only possible to be sure of the correct value of the integer when it was not greater than two or three. Owing principally to the evaporation of the drops it was not possible to obtain accurate results. The results obtained varied from 2×10^{-10} to $4{\cdot}1 \times 10^{-10}$ electrostatic units. The mean result was $3{\cdot}1 \times 10^{-10}$.

17. Millikan's Method.

The problem of making an accurate direct determination of the ionic charge was finally solved by Millikan about 1908. His method was in principle the same as that used by the writer, but he eliminated

the error due to evaporation by using small drops of oil or mercury, which do not evaporate, and by using a strong electric field and drops carrying several ionic charges he was able to make the drops move up or down and so keep a single drop under observation for a long time and make a series of measurements with it. Millikan made a long series of measurements, taking all possible precautions to eliminate errors, and his final result is believed to be correct to one part in one thousand.

Millikan's final form of apparatus is shown diagrammatically in fig. 6. AA and BB are two circular metal plates with optically worked plane surfaces. These plates are separated by three glass blocks cut from a piece of plane parallel optically worked glass. The distance between the plates is 14·9174 mm. At the centre of the upper plate there is a very small hole as shown. The tube T leads to a sprayer S by means of which oil or mercury spray can be produced and blown into the space above the plates. The plates are supported inside a metal box CC immersed in an oil tank DD. Glass windows WW enable a beam of light to be passed between the plates so as to illuminate the drops at the centre of the plates. The drops were observed through a telescope in a direction slightly inclined to the beam of light. X-rays could also be passed between the plates when desired. The plates could be connected to a battery giving up to 10,000 volts.

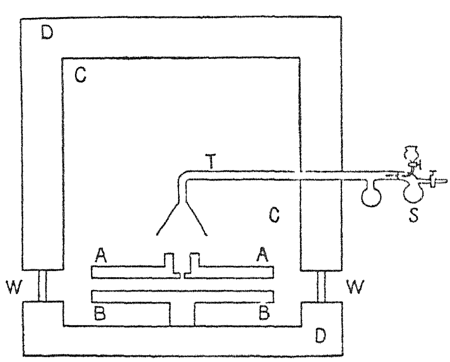


Fig. 6

On working the sprayer for a few seconds a large number of drops is formed above the plates, and a few of these fall through the small hole into the space between the plates. The drops are charged, and usually a few have such masses and charges that they remain suspended in the electric field or move very slowly up or down. A drop which moves slowly up may be selected, and its small upward velocity is determined by measuring with a chronograph the time the image of the drop seen in the telescope takes to go from one cross hair in the eyepiece to another. The images of the drops are seen as minute points of light like stars, and the instant at which they pass a cross hair can be very accurately timed. The electric field is then cut off and the velocity with which the drop falls is determined. The field is then put on and the upward velocity again found. In this way a single drop can be kept under observation for several hours and a

great many observations made on it. The charge on the drop can be varied when desired by passing X-rays between the plates. The air is ionized by the rays and an ion usually gets on to the drop very soon. The change in the charge on the drop when an ion gets on it is immediately made evident by the change in the velocity of the drop in the electric field.

For a particular drop the charge on it is proportional to $v_2 - v_1$, the difference between its velocity in the field and that due to gravity alone. When v_2 is an upward velocity the charge is proportional to the sum of the numerical values of the two velocities. The following table gives a set of numbers equal to $v_2 - v_1$ multiplied by a constant (A) which were obtained by Millikan in a series of observations on one oil drop. The charge on this drop was changed from time to time and its velocities were found.

$A(v_2 - v_1)$			$A(v_2 - v_1)$		
19·66 ÷	4	4·915	53·91	11	4·901
24·60 ÷	5	4·920	59·12	12	4·927
29·62	6	4·937	63·68	13	4·898
34·47	7	4·924	68·65	14	4·904
39·38	8	4·922	78·34 ÷	16	4·896
44·42 ÷	9	4·935	83·22 ÷	17	4·895
49·47	10	4·947			

All the numbers found for $A(v_2 - v_1)$ when divided by an integer give nearly equal numbers. This shows that the charges on the drop were always exact multiples of a definite unit, which was the ionic charge, because the charge was varied by adding positive or negative ions one at a time. These results show that all the ions carry the same charge either positive or negative, so that the ionic charge is not an average value for a quantity which varies, but a definite atomic unit.

The absolute value of e was found from a very large number of measurements on many different drops. The viscosity of dry air was very carefully redetermined, and also of course the exact density of the oil used.

It was found that Stokes's law for a falling drop,

$$v = \frac{2}{9} \frac{\rho - \rho'}{\mu} g a^2,$$

is not exactly true for very small drops, and the deviation was accurately determined and allowed for. The final result obtained was

$$e = 4{\cdot}80 \times 10^{-10} \text{ electrostatic units.}$$

The ionic charge has been determined by several other methods in recent years. Rutherford and Geiger determined the charge $2e$ on α-rays and got $e = 4{\cdot}65 \times 10^{-10}$ and Regener by the same method

got $4{\cdot}79 \times 10^{-10}$. This method is described in the chapter on Cathode Rays, β-rays, and α-rays.

18. Perrin's Investigations. Brownian Movements and Diffusion.

Perrin has made a very interesting series of investigations on the Brownian movements of small particles suspended in liquids, from which he has deduced the number of molecules in unit volume of any gas and hence the value of the ionic charge. Very small particles suspended in a liquid continually move about in an irregular manner visible in a high-power microscope. This motion is believed to be due to collisions between the particles and the molecules of the liquid. The particle may be regarded as a big molecule, and its average kinetic energy should be equal to the average kinetic energy of a gas molecule at the same temperature.

An emulsion containing n equal particles per unit volume suspended in a liquid may be compared with a gas containing n molecules in unit volume. The particles may be regarded as exerting a pressure equal to $\frac{1}{3}mnV^2$, where m is the mass of a particle, and V^2 is the average value of the square of the velocities of the particles. If we consider a horizontal layer of thickness dx in the emulsion, the downward force of gravity on the ndx particles in unit area of the layer must be balanced by the upward force arising from the variation of the gas pressure p due to the motion of the particles. If ρ' is the density of the liquid and ρ that of the particles, we have therefore

$$\frac{nmg(\rho-\rho')}{\rho}\,dx = -\frac{\partial p}{\partial x}\,dx.$$

Putting $p = \frac{1}{3}mnV^2$, we get

$$\frac{1}{n}\frac{\partial n}{\partial x} = -3g\frac{\rho-\rho'}{\rho V^2}.$$

Integrating this, and putting $n = n_0$ at $x = 0$, we find

$$\log\frac{n_0}{n} = 3g\frac{\rho-\rho'}{\rho V^2}\,x.$$

Perrin prepared an emulsion of equal gamboge particles and placed a small quantity of it in a shallow glass cell. The particles in it were observed with a vertical microscope focused on a horizontal plane in the emulsion. The number of particles visible in the focal plane was counted, and by moving the microscope up and down through known distances the variation of this number with the level of the focal plane was determined. This gave $\frac{1}{x}\log\frac{n_0}{n}$ and so V^2 could be calculated by means of the above equation. The densities ρ and ρ' and the size of the particles were also determined, so that the mass m and hence mV^2

for the particles could be calculated. Now [illegible] is equal to $3p/n$ where p is the pressure of any gas containing n molecules in unit volume. Thus Perrin was able to calculate the number of molecules in unit volume of any gas at 760 mm. pressure and at the temperature of his experiments. From this the number N of molecules in a gramme-molecule or mol immediately follows, and since $Ne = 9650$ the ionic charge e can be deduced. In this way Perrin got $e = 4{\cdot}2 \times 10^{-10}$ electrostatic units.

Perrin also observed the diffusion of the particles through the liquid. If n is the number of particles in unit volume and K the coefficient of diffusion we have

$$-K\frac{\partial n}{\partial x} = un,$$

where u is the velocity of diffusion in the x direction. If v and w are the velocity components due to diffusion in the y and z directions we have also

$$-K\frac{\partial n}{\partial y} = vn,$$

$$-K\frac{\partial n}{\partial z} = wn.$$

But $$\frac{\partial}{\partial x}(un) + \frac{\partial}{\partial y}(vn) + \frac{\partial}{\partial z}(wn) = -\frac{\partial n}{\partial t},$$

so that $$K\left(\frac{\partial^2 n}{\partial x^2} + \frac{\partial^2 n}{\partial y^2} + \frac{\partial^2 n}{\partial z^2}\right) = \frac{\partial n}{\partial t}.$$

Now consider a large number N of particles distributed in any manner over a certain region in a liquid, and let them all be very far from the boundary of the liquid. The particles will diffuse about so that the region they occupy will gradually get larger. Consider the mean value of the square of the distance R of a particle from any fixed point, and take this point as the origin, so that $R^2 = x^2 + y^2 + z^2$. The mean value of R^2 is given by

$$\overline{R^2} = \frac{\int (x^2 + y^2 + z^2)\, n\, dx\, dy\, dz}{\int n\, dx\, dy\, dz},$$

where the integration is extended over all the space where n is not zero. Differentiating with respect to the time t we get

$$N\frac{\partial \overline{R^2}}{\partial t} = \int (x^2 + y^2 + z^2)\frac{\partial n}{\partial t}\, dx\, dy\, dz,$$

since $N = \int n\, dx\, dy\, dz$.

Putting $K\left(\frac{\partial^2 n}{\partial x^2}+\frac{\partial^2 n}{\partial y^2}+\frac{\partial^2 n}{\partial z^2}\right)$ for $\frac{\partial n}{\partial t}$, we obtain

$$N\frac{dR^2}{dt} = K\int (x^2+y^2+z^2)\left(\frac{\partial^2 n}{\partial x^2}+\frac{\partial^2 n}{\partial y^2}+\frac{\partial^2 n}{\partial z^2}\right)dx\,dy\,dz$$

$$= K\int\left(x^2\frac{\partial^2 n}{\partial x^2}+y^2\frac{\partial^2 n}{\partial y^2}+z^2\frac{\partial^2 n}{\partial z^2}\right)dx\,dy\,dz +$$

$$K\int\left\{\frac{\partial}{\partial x}\left((y^2+z^2)\frac{\partial n}{\partial x}\right)+\frac{\partial}{\partial y}\left((x^2+z^2)\frac{\partial n}{\partial y}\right)+\frac{\partial}{\partial z}\left((x^2+y^2)\frac{\partial n}{\partial z}\right)\right\}dx\,dy\,dz.$$

The second integral may be transformed into a surface integral by Green's theorem, and, since it is to be taken over a volume so large that n is zero near the surface enclosing the volume, this integral is zero.

Integrating $\int x^2\frac{\partial^2 n}{\partial x^2}dx$ by parts gives

$$\left[x^2\frac{\partial n}{\partial x}\right] - 2\int\frac{\partial n}{\partial x}x\,dx = -2\int x\,dn,$$

since $\left[x^2\frac{\partial n}{\partial x}\right]$ is zero at both limits.

Again $$-2\int x\,dn = -2[nx] + 2\int n\,dx = 2\int n\,dx,$$

since $[nx]$ is zero at the limits.

In the same way $$\int y^2\frac{\partial^2 n}{\partial y^2}dy = 2\int n\,dy,$$

and $$\int z^2\frac{\partial^2 n}{\partial z^2}dz = 2\int n\,dz.$$

Hence $$K\int\left(x^2\frac{\partial^2 n}{\partial x^2}+y^2\frac{\partial^2 n}{\partial y^2}+z^2\frac{\partial^2 n}{\partial z^2}\right)dx\,dy\,dz = 6K\int n\,dx\,dy\,dz,$$

so that we get $$\frac{dR^2}{dt} = 6K.$$

Since $R^2 = x^2+y^2+z^2$, we see that the mean square of the distances of the particles from a straight line will increase with the time at the rate $4K$, and the mean square of the distances from a plane at the rate $2K$. The equation defining the coefficient of diffusion K,

$$-K\frac{\partial n}{\partial x} = nu,$$

gives, since the gas pressure p due to the particles is proportional to n,

$$-K\frac{\partial p}{\partial x} = pu = \tfrac{1}{3}mnuV^2.$$

But $-\partial p/\partial x$ may be regarded as the force on the n particles in unit volume which causes them to move with the velocity u, so that if s denotes the average velocity with which the particles fall through the liquid under the force of gravity $m\frac{\rho-\rho'}{\rho}g$, we have

$$\frac{-\frac{\partial p}{\partial x}}{u}=\frac{nm\frac{\rho-\rho'}{\rho}g}{s}=\frac{1}{3}\frac{mnV^2}{K},$$

or

$$V^2=\frac{3Kg(\rho-\rho')}{s\rho}.$$

This equation enables V^2 to be calculated from K and s. Perrin determined K by observing the positions of a number of the particles at known times and so getting the rate at which the mean square of their distances from a vertical plane increased with the time. This rate, as we have just seen, is equal to $2K$. He also found s by allowing a group of the particles to fall through the liquid for a known time. In this way he got an independent estimate of V^2 and so of the ionic charge. This estimate agreed with his previous result got from the distribution of the particles with respect to height in the liquid.

Perrin's results provide an interesting confirmation of the atomic theory of matter and of electricity. His value of e is generally regarded as less reliable than that obtained by Millikan.

The value of the ionic charge was deduced by Planck from his theory of heat radiation. This is discussed in the chapter on the Quantum Theory.

REFERENCES

1. *Electricity in Gases*. J. S. Townsend.
2. *The Electron*. R. A. Millikan.
3. *Atoms*. J. Perrin.

CHAPTER XIV

The Motion of Electrons in Gases

1. Townsend's Apparatus.

When ions are produced in gases, by X-rays or in other ways, at atmospheric pressure, it is found that they consist of clusters of several molecules carrying the electronic charge e. At lower pressures in many cases the positive ions are clusters of molecules, but the negative ions are merely free electrons. The motion of these electrons in gases has been studied extensively by J. S. Townsend, of Oxford, and others, and this chapter is devoted mainly to a discussion of Townsend's experiments.

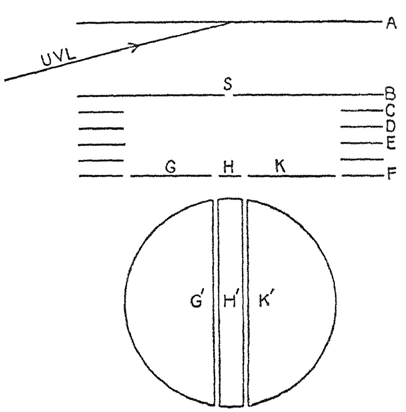


Fig. 1

The apparatus used by Townsend is shown diagrammatically in fig. 1. A and B are two parallel circular metal plates about 4 cm. apart. At the centre of B there is a slit S which is 2 mm. wide and 1·5 cm. long. C, D, E, and F are metal rings cut out of thin sheets. These rings are equally spaced about 1 cm. apart. In the plane of the ring F there is a circular disc GHK which nearly fills the hole in the ring. This disc is divided into three parts by two straight parallel cuts equally distant from the centre. The disc is shown in plan at G′H′K′. The different parts of the apparatus are supported inside a metal case not shown in the figure. The case can be evacuated and filled with gas at any desired pressure. By allowing ultra-violet light to fall on the under side of the plate A electrons can be set free at its surface. The plate B is kept at a higher potential than A so that the electrons emitted by A move through the gas towards B. Some of them pass through the slit S into the space between B and GHK. The disc GHK and the ring F are kept at a higher potential than the

plate B, so that the electrons which pass through the slit continue to move down and finally reach the disc. The rings C, D, E are used to make the electric field between S and H more nearly uniform. They are kept at potentials such that the potential differences between them and B are proportional to their distances from the plate B. The different potential differences are maintained by suitable batteries, and the disc GHK and ring F are kept at zero potential. The electric field above B is always made equal to that below, so that the electrons acquire a uniform velocity above B and continue to move with the same velocity below B. The length of the slit S is parallel to the sides of the strip H. The strip H is 1·5 mm. wide and 7 cm. long, with gaps on each side of it 0·5 mm. wide. The distance between the plate B and the disc GHK is 1 cm. As the electrons move down from the slit towards H they diffuse out sideways so that some of them fall on H and some on G and K. The three parts of the disc can be insulated and can be connected to a quadrant electrometer and induction balance by means of which the negative charges which they receive can be measured. Let n_1, n_2, and n_3 be the charges received by G, H, and K respectively. The sideways diffusion of the electrons then depends on the ratio $R = \dfrac{n_2}{n_1 + n_2 + n_3}$. This ratio was determined for a number of gas pressures p, and values of the electric field strength Z along the path of the electrons.

2. Mathematical Theory of Townsend's Experiment.

Let us now consider the theory of this experiment. If K is the coefficient of diffusion of the electrons in the gas, then, using rectangular axes x, y, and z, we have $-K\dfrac{\partial n}{\partial x} = nv_x$, where n is the number of electrons per unit volume, and v_x their average velocity component along x.

Also

$$-K\frac{\partial n}{\partial y} = nv_y,$$

$$-K\frac{\partial n}{\partial z} = nv_z.$$

The partial pressure P of the electrons is proportional to n, so that

$$-K\frac{\partial P}{\partial x} = Pv_x,$$

with similar equations in y and z. For a steady state, in which $\dfrac{\partial n}{\partial t} = 0$ everywhere, we have div $(n\mathbf{v}) = 0$, or div $(P\mathbf{v}) = 0$, since $n\mathbf{v}$ is the number of electrons flowing through unit area in unit time, and no new electrons are generated in the gas. The equation $-\dfrac{\partial P}{\partial z} = \dfrac{Pv_z}{K}$ may be said to give the velocity of the electrons

due to a force $\frac{\partial P}{\partial z}$ on the electrons in unit volume. If there is present an electric field Z along the z axis, this gives an additional force neZ on the electrons, so that

$$neZ - \frac{\partial P}{\partial z} = \frac{Pv_z}{K}.$$

In Townsend's apparatus take the origin at the centre of the slit and the z-axis vertically downwards along the path of the electrons. Also take y along the slit, and x perpendicular to the slit in the plane of the plate B.

The equations of motion of the electrons are then,

$$-\frac{\partial P}{\partial x} = \frac{Pv_x}{K},$$

$$-\frac{\partial P}{\partial y} = \frac{Pv_y}{K},$$

$$neZ - \frac{\partial P}{\partial z} = \frac{Pv_z}{K}.$$

Differentiating these with respect to x, y, z respectively, and using div $(P\mathbf{v}) = 0$,

we get $$\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} + \frac{\partial^2 P}{\partial z^2} = Ze\frac{\partial n}{\partial z}.$$

The distribution of the electrons between the three parts of the disc is determined almost entirely by the diffusion along the x axis, so that for an approximate solution we may use $\frac{\partial^2 P}{\partial x^2} = Ze\frac{\partial n}{\partial z}$. This would be exactly correct if the slit were infinitely long. For a gas at temperature T we have $P = nkT$, where k is the gas constant for one molecule. It is found that the average kinetic energy of the electrons exceeds that of gas molecules, so that the temperature of the electrons is greater than the temperature of the gas through which they are moving. For the electrons, then, let $P = \beta nkT$, where T is the temperature of the gas and βT that of the electrons. This gives

$$\frac{\partial^2 n}{\partial x^2} = \frac{Ze}{\beta kT}\frac{\partial n}{\partial z}.$$

The factor e/kT is the same as Ne/NkT, where N is the number of molecules in one gramme-molecule of any gas. We have

$$Ne = 9650 \text{ electromagnetic units},$$
$$Nk = 8{\cdot}315 \times 10^7,$$

so that, with $T = 273 + 15$, which was about the temperature in all Townsend's work, we get

$$\frac{\partial^2 n}{\partial x^2} = 40{\cdot}3\frac{Z}{\beta}\frac{\partial n}{\partial z},$$

where Z is now expressed in volts per centimetre. A solution of this equation is

$$n = Az^{-1/2}e^{-\alpha x^2/z},$$

where A and α are constants. This makes $n = 0$ at $z = 0$, except where $x = 0$, and so approximately satisfies the conditions in Townsend's apparatus; for, at

$z = 0$, n is zero, except on the slit where x is nearly zero. On substituting, we find $\alpha = 10{\cdot}1 Z/\beta$. The ratio R is therefore given by

$$R = \frac{\int_{-b/2}^{b/2} e^{-\alpha x^2/4z}\,dx}{\int_{-\infty}^{\infty} e^{-\alpha x^2/4z}\,dx} = \frac{2}{\sqrt{\pi}} \int_0^{0{\cdot}1\sqrt{Z/\beta}} e^{-y^2}\,dy,$$

since the breadth b of the strip H, plus half the total width of the gaps on each side of it, is 0·50 cm., and $z = 4$ cm. at the disc. Townsend obtained a similar but more accurate solution of the problem, and so was able to find β from the experimentally found values of R. The above expression for R agrees nearly with Townsend's values.

It is found in this way that β is a function of Z/p in any given gas. That is, if the pressure p is changed and Z, the electric field strength, is also changed so as to keep Z/p the same, then β remains unaltered.

The following table gives some of the values of β (= temperature of electrons/temperature of gas—see above) found in nitrogen gas at 15° C.

Z/p	β
60	126·0
40	89·0
20	59·5
10	48·5
5	41·3
2	30·5
1	21·5
0·5	13·0
0·25	7·5

Z is in volts per centimetre and p in millimetres of mercury.

Thus, for example, in an electric field of 20 volts per centimetre in nitrogen at 1 mm. pressure $Z/p = 20$, and so $\beta = 59{\cdot}5$. That is, the kinetic energy of an electron is 59·5 times that of a gas molecule at 15° C. With $Z = 100$ volts per centimetre and $p = 5$ mm., β is still 59·5.

3. Average Velocity and Kinetic Energy of the Electrons.

The average velocity of the electrons in the direction of the electric field Z was found with the same apparatus by deflecting the stream of electrons by means of a magnetic field. A uniform magnetic field of strength H was produced by means of suitable coils of wire through which a current could be passed. This field was perpendicular to the electric field Z, and parallel to the length of the slit. The stream of electrons was deflected sideways by the magnetic field, and the strength of the field was adjusted until one-half of the electrons was received by the electrode G and the other half by the electrodes H and K connected together. The centre of the stream was then deflected through 2·75 mm. at the plane GHK. By reversing the field the electrons could be deflected in the opposite direction so that half fell on K and half on G and H together. If the average velocity of the electrons along the

direction of the field Z is W, then the force on n electrons due to the magnetic field H is $nHeW$ and the force due to the electric field is neZ. Consequently, if a is the deflection of the electrons while they move along the electric field a distance b, we have $\frac{HW}{Z} = \frac{a}{b}$. In Townsend's apparatus, a was 2·75 mm. when b was 40 mm., so that

$$W = \frac{Z}{H} \times 0{\cdot}06875.$$

In this way W was found for different values of Z and p.

The average kinetic energy of the electrons is β times that of a gas molecule at the same temperature. The average energy of a gas molecule is $\frac{3}{2}kT$, where k is the gas constant for one molecule, and T is the absolute temperature. If m is the mass of one electron, and V^2 the average of the squares of the velocities of the electrons, then

$$\tfrac{3}{2}\beta kT = \tfrac{1}{2}mV^2,$$

so that
$$V^2 = \frac{3\beta kT}{m}.$$

As the electrons move through the gas they acquire energy by the action of the electric field and lose it to the gas molecules by collisions. The velocity V with which the electrons move about is continually changed in direction by collisions, and it is much greater than W, the average velocity of drift along the electric field. V is usually called the velocity of agitation. It is easy to calculate W approximately. The rate at which n electrons receive momentum in the z direction owing to the electric field is neZ. The momentum of these electrons in the z direction when a steady state has been reached is nmW. When an electron collides with a gas molecule, we shall assume that after the collision it is as likely to be moving in one direction as another. Consequently the momentum of n electrons immediately after collisions is zero. On the average, then, an electron loses momentum mW per collision. The number of collisions made by n electrons in unit time is approximately nV/λ, where λ is the mean free path. Hence the momentum lost by n electrons in unit time is $\frac{nV}{\lambda} \times mW$, so that in a steady state we have

$$neZ = \frac{nmVW}{\lambda},$$

and therefore
$$\lambda = \frac{mVW}{Ze}.$$

Townsend's experiments give V and W, so that λ can be calculated.

The energy received by n electrons in unit time from the electric field is $neZW$, so that the average energy lost at a collision is $\frac{neZW}{nV/\lambda}$. Dividing this by $\frac{1}{2}mV^2$, we get the average fraction of the energy of the electron lost per collision. This is

$$\frac{2neZW\lambda}{nVmV^2} = \frac{2W^2}{V^2}, \text{ since } \frac{Ze\lambda}{mV} = W,$$

and so can be calculated.

The tables below give some of the results obtained in several gases, the symbols having the following meaning: Z = electric field, p = gas pressure; β = ratio of average energy of electron and gas molecule, W = average velocity of drift of electrons along electric field, V = their velocity of agitation (see above); f = ratio $2W^2/V^2$. Z/p is in volts per centimetre per millimetre pressure, W and V are in centimetres per second, and λ is the mean free path in centimetres at 1 mm. pressure.

Z/p	β	W	V	λ	f
		Nitrogen.			
60·0	126	193 × 10^5	129 × 10^6	0·0289	0·055
20·0	60	86 ,,	89 ,,	0·0266	0·0234
5·0	41	27 ,,	74 ,,	0·0277	0·0033
1·0	22	9 ,,	54 ,,	0·032	0·0007
0·25	8	5 ,,	32 ,,	0·045	0·0007
		Hydrogen.			
50·0	148	217 × 10^5	140 × 10^6	0·012	0·059
20·0	78	70 ,,	102 ,,	0·025	0·0117
5·0	26	26 ,,	59 ,,	0·021	0·0046
1·0	9	12 ,,	35 ,,	0·029	0·0029
0·25	3	7 ,,	20 ,,	0·036	0·0026
		Argon.			
15·0	324	82 × 10^5	207 × 10^6	0·079	0·000386
5·0	310	40 ,,	202 ,,	0·113	0·000097
0·95	280	6 ,,	193 ,,	0·085	0·000024
0·195	120	3 ,,	126 ,,	0·147	0·000016
		Helium.			
5·0	172	30 × 10^5	151 × 10^6	0·064	0·00098
1·0	53	8·3 ,,	81 ,,	0·049	0·00024
0·1	6·2	3 ,,	29 ,,	0·06	0·00026
0·013	1·77	1·1 ,,	15 ,,	0·091	0·00013

The velocity of drift W is not proportional to Z/p; it increases rather mor[e] rapidly than $\sqrt{Z/p}$ in most cases. This is due mainly to the increase in the velocit[y]

of agitation In the inactive gases, argon and helium, the energy lost at collisions is very small, showing that the collisions are almost perfectly elastic; even in nitrogen the loss is only about one part in one thousand when Z/p is small.

The mean free path varies with Z/p, and seems to pass through a minimum value as Z/p increases The mean free path in argon is surprisingly large when Z/p is small.

4 **Ionization by Collisions.**

When a stream of electrons is passing through a gas in an electric field, and the kinetic energy of the electrons is great enough, some of the collisions between electrons and molecules cause an electron to escape from the molecule. A molecule is said to be ionized by such a collision, since a molecule which has lost an electron forms a positive ion The ionization of gases by a stream of rapidly moving electrons was observed by Lenard, who passed Lenard rays, which are rapidly moving electrons, through different gases. The ionization of gases by collisions has been investigated by Townsend, and his results will be discussed here. Other important investigations on ionization by collisions are described in the chapter on ionization and radiation potentials. In Townsend's experiments electrons were set free at the surface of a metal plate by means of ultra-violet light. A uniform electric field was maintained between this plate and another parallel plate, which was charged positively, so that the electrons moved across from the first plate to the second. The charge received by the second plate was measured, and of course is proportional to the number of electrons which arrive at it. The space between the two plates was filled with a gas at pressure p, and the distance between the two plates was varied, keeping the electric field strength between them constant by making the potential difference proportional to the distance between the plates. It was found that the charge received by the second plate increased with the distance d between the plates, so that $q = q_0\epsilon^{ad}$, where q is the charge at distance d, q_0 the value of q when d is very small, and a a constant.

Let us suppose that n_0 electrons are set free by the ultra-violet light and that as these move across to the opposite plate they ionize the gas molecules, each setting free a electrons from the gas molecules while moving 1 cm. in the direction of the electric field. Then if n electrons pass through a plane at a distance x from the plate from which n_0 start, we have $dn = an\,dx$. This gives $\log n = ax +$ constant, so that since $n = n_0$ at $x = 0$ we get

$$n = n_0\epsilon^{ax}.$$

Thus, putting $q = ne$ and $q_0 = n_0e$, we get $q = q_0\epsilon^{ad}$, as was found experimentally. By measuring q and q_0 Townsend determined the value of a for different electric field strengths X, and different gas pressures p. It was found that a/p is a function of X/p for a given

gas at a constant temperature, so that $\alpha = pf(X/p)$. This means, for example, that if X and p are both increased n times so that X/p remains the same, then α is also increased n times. We have seen that the kinetic energy of the electrons is a function of X/p, and the same is true of the velocity W with which they move through the gas in the direction of the electric field. We should expect the number of molecules ionized to depend on the kinetic energy of the electrons, and to be proportional to the number of collisions between molecules and electrons. Suppose n electrons pass through a layer of the gas of thickness dx. The time each electron is in this layer is dx/W, and the n electrons make nV/λ collisions in unit time, V being the velocity of agitation and λ the mean free path, as before. The number of collisions in the layer is, therefore,

$$\frac{dx}{W}\frac{nV}{\lambda}.$$

If a fraction F of the collisions result in ionization, the number of electrons liberated per unit length is $FnV/\lambda W$, so that

$$\alpha = FV/\lambda W.$$

Now V and W are both functions of X/p, λ is inversely as p, and F must depend on $\frac{1}{2}mV^2$ or V only, and so is also a function of X/p, hence we must have $\alpha = pf(X/p)$, in agreement with the experimental results. If we suppose that all collisions for which the velocity of the electron is greater than a certain value result in ionization then we can calculate the fraction F. Let V now denote the velocity of agitation of an individual electron, and V^2 the average of the squares of the velocities of the electrons, and assume that the distribution of the velocities V is given by Maxwell's law; that is, let the number of electrons in unit volume for which V is between V and $V+dV$ be $AV^2\epsilon^{-qV^2}dV$, so that the number of collisions made by these electrons in unit time is proportional to $V^3\epsilon^{-qV^2}dV$. The constant q will be equal to $\frac{m}{2\beta kT}$, where m is the mass of an electron and k is the gas constant for one molecule.

The number of collisions for which V is greater than V_0, divided by the total number, is therefore given by

$$F = \frac{\int_{V_0}^{\infty} V^3\epsilon^{-qV^2}dV}{\int_0^{\infty} V^3\epsilon^{-qV^2}dV},$$

Hence
$$F = (qV_0^2 + 1)\epsilon^{-qV_0^2},$$

so that
$$\alpha = \frac{V}{\lambda W}(qV_0^2 + 1)\epsilon^{-qV_0^2}.$$

We have $\lambda = \lambda_1/p$, where λ_1 is the mean free path at 1 mm. pressure, and also $p = \frac{1}{3}mnV^2 = \beta nkT$, so that $\beta kT = \frac{1}{3}m\overline{V}^2$, and $q = \frac{m}{2\beta kT}$.

Hence
$$qV_0^2 = \frac{3}{2}\frac{V_0^2}{\overline{V}^2},$$

and
$$\alpha = \frac{p}{\lambda_1 W}\left(\frac{3}{2}\frac{V_0^2}{V} + V\right)\epsilon^{-3V_0^2/2\overline{V}^2}.$$

When $\overline{V}/V_0$ is large, this becomes approximately $\alpha = pV/\lambda_1 W$, so that if V and W become equal when V is large, $\alpha = \frac{p}{\lambda_1} = \frac{1}{\lambda}$. This theory indicates that α should be very small when V is much less than V_0 and should increase rapidly with V when V_0 and V are about equal. For large values of V the rate of increase of α with V should become small. The assumption that all collisions for which V is greater than V_0 result in ionization is not really correct, and it is found that the very fast electrons of cathode rays produce fewer ions than slower electrons, so that this theory is not satisfactory and is merely intended as a rough illustration. It is not possible to deduce from measurements of α any accurate estimate of the velocity which an electron must have to be able to ionize a molecule. Accurate methods of measuring this velocity have been developed and are described in the chapter on ionization and radiation potentials.

The following table gives some of the values of α/p found by Townsend in different gases for the values of X/p given in the first row of the table.

X/p	1000.	800	600	400	200.	100.
Air	10 5	9 3	7·9	5·82	2 6	0·72
Water vapour	9 7	9 0	7 95	6·35	3·6	1 31
Hydrogen ..	—	-		3 7	2·62	1·36
Argon	—		9 2	7·5	4 4	2 0
Helium	-	-			2 37	2·0

5. Ionization by Positive Ions.

The equation $q = q_0\epsilon^{\alpha d}$, which was found to give the charge q received by an electrode at a distance d from the plate at which electrons carrying a charge q_0 start, gives q accurately when α is small and d not too large. However, it was found, with large values of X/p, for which α is large, that when d also was large the charge q obtained was greater than that given by this equation. This was explained by Townsend by means of the hypothesis that the positive ions also

ionize molecules by collisions when X/p is large. Another possible explanation is that the positive ions liberate electrons from the negatively charged electrode when they strike it. The results obtained can be explained equally well in either way, and Townsend's experiments do not enable us to decide which explanation is correct. We shall consider Townsend's explanation here, and the alternative one in section 7.

Let a positive ion ionize β molecules while it moves unit distance in the direction of the electric field X in gas at pressure p. Let n_0 electrons start from an electrode at $x = 0$, and let n be the number which pass through a plane at a distance x from this electrode. Also let m be the number of positive ions which pass through this plane in the opposite direction. Then between the planes at x and $x + dx$ the number of molecules ionized is

$$n\alpha\,dx + m\beta\,dx,$$

so that
$$dn = n\alpha\,dx + m\beta\,dx,$$

and
$$dm = -n\alpha\,dx - m\beta\,dx,$$

since each molecule ionized gives an electron and a positive ion. This gives $dn + dm = 0$, so that $n + m =$ constant. At $x = 0$ we have $n = n_0$, and if d is the distance between the electrodes then at $x = d$ we have $m = 0$, so that the constant is equal to n_d. Hence

$$n + m = n_d,$$

and
$$dn = n\alpha\,dx + (n_d - n)\beta\,dx = n(\alpha - \beta)\,dx + n_d\beta\,dx.$$

Integrating this, we get

$$\frac{1}{\alpha - \beta}\log\{n(\alpha - \beta) + n_d\beta\} = x + \text{constant}.$$

At $x = 0$, $n = n_0$, so that

$$\frac{1}{\alpha - \beta}\log\{n_0(\alpha - \beta) + n_d\beta\} = \text{constant}.$$

Hence
$$\frac{1}{\alpha - \beta}\log\frac{n(\alpha - \beta) + n_d\beta}{n_0(\alpha - \beta) + n_d\beta} = x.$$

At $x = d$, $n = n_d$, so that

$$\frac{1}{\alpha - \beta}\log\frac{n_d\alpha}{n_0(\alpha - \beta) + n_d\beta} = d,$$

or
$$n_d = \frac{n_0(\alpha - \beta)}{\alpha\epsilon^{-d(\alpha - \beta)} - \beta}.$$

When $\beta = 0$, this reduces to

$$n_d = n_0 \epsilon^{ad},$$

as we should expect. By measuring $q = n_d e$ and $q_0 = n_0 e$ for different values of d, the values of α and β can be determined When d is small, $n_d = n_0 \epsilon^{ad}$ nearly, which gives an approximate value of α. The best values of α and β can then be found by trial.

The following table contains some results given by Townsend for air at 4 mm. pressure with an electric field of 700 volts per centimetre. The value of α used was 8·16 and that of β was 0 0067.

d.	q/q_0.	ϵ^{ad}.	$\dfrac{\alpha - \beta}{\alpha\epsilon^{-d(\alpha-\beta)} - \beta}$.
Cm			
0 2	5 12	4·9	5 11
0 3	11 4	11 6	11 6
0 4	26 7	26 1	26 5
0 5	61 0	59 0	62 0
0 6	148 0	133 0	149 0
0 7	401 0	301 0	399 0
0 8	1500 0	680·0	1544 0

We see that ϵ^{ad} represents the values of q/q_0 quite well up to $d = 0·5$ cm., but for larger values of d, ϵ^{ad} is too small. The expression $\dfrac{\alpha - \beta}{\alpha\epsilon^{-d(\alpha-\beta)} - \beta}$, however, agrees well with all the values of q. Similar results were obtained by Townsend with air and other gases at different pressures, and with different electric fields. It was found that $\beta = p\phi\left(\dfrac{X}{p}\right)$, where $\phi\left(\dfrac{X}{p}\right)$ denotes a function of X/p. β is always much smaller than α. The following table gives some values of β/p.

X/p.	Hydrogen.	Argon.	Air.	Carbon Dioxide.
200	0 08	0·02	--	-
600	--	0 22	0·10	--
800	--	--	0 18	0 01
1400	--	---	—	0·10

6. **Sparking Potentials.**

The equation

$$\frac{q}{q_0} = \frac{\alpha - \beta}{\alpha\epsilon^{-d(\alpha-\beta)} - \beta}$$

shows that q/q_0 becomes infinite when $\alpha = \beta\epsilon^{d(\alpha-\beta)}$. This means that

a discharge once started would continue indefinitely. The equation $\alpha = \beta\epsilon^{d(\alpha-\beta)}$ therefore gives the sparking potential for the distance d, pressure p, and field strength X. The sparking potential S is given by $S = Xd$. When q/q_0 is infinite a single electron set free at the negative electrode is sufficient to start a continuous electric discharge. Now the gas always contains a few free electrons owing to the presence of traces of radioactive bodies, so that if the potential difference between the two parallel plates is gradually increased until the field reaches the value for which $\alpha = \beta\epsilon^{d(\alpha-\beta)}$, a continuous discharge will then start in the gas between the plates.

Now $\frac{X'}{p} = \frac{S}{pd}$, where X' is the field strength for the sparking potential S. Hence we have $\alpha = pf(S/pd)$, and $\beta = p\phi(S/pd)$, where $f(S/pd)$ and $\phi(S/pd)$ denote functions of S/pd. Hence the equation $\alpha = \beta\epsilon^{d(\alpha-\beta)}$ may be written

$$f(S/pd) = \phi(S/pd)\epsilon^{pd[f(S/pd)-\phi(S/pd)]}.$$

This is a relation between S and pd, so that we must have $S = S(pd)$, where $S(pd)$ denotes a function of pd only. It is found experimentally that the sparking potential between parallel plates in any gas depends on the product of the gas pressure and the distance between the plates. Townsend found that the observed sparking potentials agreed well with those given by the equation

$$\alpha = \beta\epsilon^{d(\alpha-\beta)}.$$

For example, in air at 1 mm. pressure, with an electric field of 700 volts per centimetre, it was found that $\alpha = 8{\cdot}16$ and $\beta = 0{\cdot}0067$. These values of α and β give $d = 0{\cdot}871$ cm. when substituted in the condition for sparking. Hence $S = 0{\cdot}871 \times 700 = 609$ volts. It was found that with $d = 0{\cdot}871$ cm. a continuous discharge began when the potential was increased to 615 volts, which agrees well with 609 volts.

The sparking potential $S = S(pd)$ has a minimum for a certain value of pd. If p is kept constant and d gradually increased from a very small value, S at first falls as d increases, then passes through a minimum value and rises again. When pd is large, S is nearly proportional to pd. If d is kept constant and the gas pressure varied, when p is very small the sparking potential is very large. As p increases the sparking potential falls to the minimum value, and then rises and becomes nearly proportional to p.

The following table gives some sparking potentials S in several gases between parallel plates 3 mm. apart. The pressures p are in millimetres of mercury, and S is in volts.

Air.		Hydrogen		Sulphur Dioxide	
p	S	p	S	p	S
51 0	1480	13 6	415	13 5	1145
21 4	790	8 54	356	4 5	651
5 99	452	5 4	301	1·61	471
2 51	371	4 02	278	1·04	457
1·89	356	3·44	282	0 8	465
1 22	375	2 52	310	0·43	621
0 928	441	2 15	356	0 23	1590
0 536	863	1 35	780	--	--
0·357	1786	0·861	1789	--	—

These results are due to Carr. The two parallel electrodes were separated by a ring of ebonite 3 mm. thick, which prevented the discharge passing between the backs or the edges of the electrodes. This precaution is necessary when the pressure is below that at which S is a minimum, because then the discharge tends to pass across any available path which is longer than the shortest distance between the electrodes. The fact that in any gas the sparking potential is a function of pd was first discovered by Paschen.

7 Alternative Theory of Action of Positive Ions.

Let us now suppose that the positive ions do not ionize the gas molecules, but that they set free electrons when they strike the negative electrode. Let n_0 electrons set free by ultra-violet light start from the negative electrode and move across to the positive electrode. The number of electrons which arrive at the positive electrode is then $n_0\epsilon^{ad}$, and the number of positive ions produced is $n_0(\epsilon^{ad}-1)$. These positive ions strike the negative electrode, and we suppose that each on the average sets free γ electrons at it. Hence $\gamma n_0(\epsilon^{ad}-1)$ are set free and move across like the original n_0 electrons, and so on. The total number of electrons that finally get to the positive electrode is therefore

$$n = n_0\epsilon^{ad}\{1+\gamma(\epsilon^{ad}-1)+\gamma^2(\epsilon^{ad}-1)^2+\ldots\}$$

$$= \frac{n_0\epsilon^{ad}}{1-\gamma(\epsilon^{ad}-1)}.$$

If we put $\gamma = \dfrac{\beta'}{a-\beta'}$, we get

$$n = \frac{n_0(a-\beta')\epsilon^{ad}}{a-\beta'\epsilon^{ad}}.$$

Comparing this with the equation previously obtained on the first theory, viz.

$$n = \frac{n_0(a-\beta)\epsilon^{d(a-\beta)}}{a-\beta\epsilon^{d(a-\beta)}},$$

we see that since β is always very small compared with α the two equations are practically identical. It follows that Townsend's experiments may be explained equally well by supposing that the positive ions only act by liberating electrons at the negative electrode. It has been shown experimentally that positive ions do set free electrons when they strike a metallic surface in a vacuum, but further experiments are necessary to decide on the relative importance of the two effects under different conditions. Some physicists consider that positive ions do not ionize molecules by collisions under any circumstances.

REFERENCE

Electricity in Gases. J. S. Townsend.

CHAPTER XV

The Electrical Conductivity of Flames

1. Conductivity of a Bunsen Flame.

The fact that flames and the gases coming from them conduct electricity was known to Faraday and has been the subject of many investigations since his time. The ionic theory of conductivity was first applied to gases by Giese as an explanation of his experiments on the conductivity of flame gases.

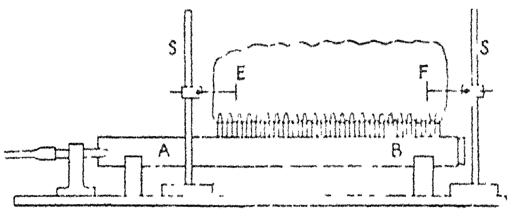


Fig. 1

The electrical conductivity of a Bunsen flame can be conveniently studied with the apparatus shown in fig. 1. A brass tube AB is supported horizontally on a wooden stand. The end near B is closed and coal gas is passed into the other end of the tube from a jet as shown. Small quartz tubes are cemented into the brass tube in a row about 25 cm. long, and the mixture of gas and air formed in the tube is burned at the ends of these quartz tubes. In this way a Bunsen flame about 10 cm. high and 26 cm. long is obtained. The quartz tubes serve to insulate the flame, so that its conductivity can be measured between two platinum electrodes E and F supported by movable stands S and S'. The electrodes may be made of pieces of sheet platinum about $1{\cdot}5 \times 1{\cdot}5$ cm. welded to stout platinum wires. If the electrodes are connected to a battery through a galvanometer or micro ammeter the current through the flame can be measured and the way it varies with the potential difference and distance between the electrodes observed.

It is found that the relation between the current C, the potential difference V, and the distance d between the electrodes is approximately

$$V = ACd + BC^2,$$

where A and B are constants, for any particular flame and electrodes. When d is small, say 1 or 2 mm., the term ACd is negligible, and the

current is nearly proportional to the square root of the potential difference.

The following table gives some values of the current in amperes obtained with a Bunsen flame similar to that just described.

Potential Difference.	Distance between Electrodes		
	1 Cm	9 Cm	18 Cm
600 volts	310 10^{-8}	295 10^{-8}	270 10^{-8}
200 ,,	175 ,,	165 ,,	113 ,,
40 ,,	67 ,,	57 ,,	18 ,,
10 ,,	22 ,,	16 ,,	13 ,,
2 ,,	5 ,,	4 ,,	3 ,,

If one of the electrodes is moved near to the surface of the flame so that it becomes cooler the current is decreased. This effect is much more marked with the negative electrode than with the positive electrode.

2. Potential Differences in the Flame.

The difference of potential between any point in the flame and one of the electrodes can easily be measured by putting in a fine insulated platinum wire and connecting it and the electrode to an electrostatic voltmeter or quadrant electrometer. The instrument used must be insulated so that the potential in the flame may not be disturbed by it. The platinum probe wire can be covered with a small fused quartz tube about $\frac{1}{2}$ mm. in diameter except for one or two millimetres at the end of the wire. In this way it is found that there is a nearly uniform potential gradient in the flame except near the electrodes. Close to the negative electrode there is usually a large and sudden drop of potential and a similar but much smaller drop close to the positive electrode. This is so when both electrodes are red hot and not near the surface of the flame. If either electrode is cooled by moving it near to the surface of the flame the drop of potential near it increases, and becomes nearly equal to the potential difference between the two electrodes if the electrode is allowed to get much cooler than the other one. The effect of cooling the negative electrode slightly is much greater than that of cooling the positive electrode

The uniform potential gradient is approximately proportional to the current. The potential difference between the electrodes may therefore be regarded as made up of three parts V_1, V_2, and V_3. V_1 is the potential drop at the positive electrode, V_2 that at the negative electrode, and V_3 that due to the uniform gradient. $V_3 = ACd$, where A is a constant, so that

$$V = ACd + V_1 + V_2.$$

Comparing this with the equation $V = ACd + BC^2$, we see that $V_1 + V_2 = BC^2$, and this result can be easily verified by measuring V_1 and V_2 with different currents passing through the flame. The layers near the electrodes in which the potential drops occur are thicker at the negative electrode than at the positive electrode. The layer at the negative electrode may be several millimetres thick while that at the positive electrode is less than 1 mm. when both electrodes are red hot.

The temperature of the hottest part of a Bunsen flame is about 2000° K. By altering the proportion of gas to air used, the temperature can be varied, and by mixing an inert gas like carbon dioxide or nitrogen with the air supplied to the flame the temperature can be considerably reduced without putting the flame out. It is found that the ratio of the current to the uniform potential gradient in the flame, which may be taken as a measure of the conductivity, increases rapidly with the temperature.

3. Ions and Electrons: Theory of Conductivity of Flames.

An ordinary Bunsen flame is a mixture of nitrogen, water vapour, carbon monoxide, and dioxide, with some hydrogen, methane, and other hydrocarbons. It is supposed that some of the molecules present in the flame become ionized by collisions with each other or with electrons or by the action of light radiation; that is, electrons are set free from some molecules so that we get positive ions and electrons in the flame. Some of the electrons may get attached to molecules, so forming negative ions. It is probable that most of the electrons set free do not get attached to molecules, at any rate in the hotter parts of the flame.

The conductivity of the flame is due to the presence of these ions and electrons. Let e denote the charge on a positive ion so that $-e$ is the charge on one electron, and let there be n_1 positive ions and n_2 electrons present in unit volume. Then, when there is an electric field in the flame, the positive ions drift along in the direction of the field and the electrons in the opposite direction. If i denotes the current density, or current per unit area of cross-section, we have

$$i = e(n_1 v_1 + n_2 v_2),$$

where v_1 is the velocity of drift of the positive ions, and v_2 that of the electrons. The velocities v_1 and v_2 are probably nearly proportional to the field strength X, at any rate when X is small. Hence $v_1 = k_1 X$ and $v_2 = k_2 X$, where k_1 and k_2 are the velocities due to unit field and are usually called the mobilities of the ions and electrons respectively. We have then

$$i = eX(n_1 k_1 + n_2 k_2).$$

The volume density of charge in the flame is equal to $e(n_1 - n_2)$, so that we have

$$\frac{dX}{dx} = 4\pi e(n_1 - n_2),$$

where x is the distance of the point considered from the positive electrode. In

the uniform gradient between the electrodes $\frac{dX}{dx} = 0$, so that $n_1 = n_2$ and therefore $i = neX_0(k_1 + k_2)$, where $n = n_1 = n_2$ and X_0 is the value of X in the uniform gradient. Now V_3, the fall of potential due to the uniform gradient, is equal to $X_0 d$, so that, since $V_3 = ACd$, we get $X_0 = AC$. If S is the area of cross-section of the flame, then, if we assume that the current is uniformly distributed over the cross-section, we have $i = C/S$, and

$$A = \frac{1}{Sne(k_1 + k_2)}.$$

Also k_2 is much larger than k_1, so that approximately $A = 1/Snek_2$.

The fact that the current is found to be nearly proportional to the uniform potential gradient shows therefore that the velocity of the electrons is nearly proportional to the field strength. This is true for fields up to 20 or 30 volts per centimetre. Near the electrodes, where the potential drops occur, dX/dx is not zero, so that n_1 is not equal to n_2.

The positive ions and the electrons attract each other so that the electrons tend to recombine with positive ions, so reforming neutral molecules. It is easy to see that this recombination must be proportional to both n_1 and n_2 or to $\alpha n_1 n_2$, where α is a constant usually called the coefficient of recombination. $\alpha n_1 n_2$ is equal to the number of neutral molecules formed per unit volume in unit time by the combination of electrons and positive ions.

Consider a layer of thickness dx between the planes at x and $x + dx$. The number of positive ions flowing into this layer per unit area per unit time through the plane at x is $n_1 v_1$, and the number flowing out through the plane at $x + dx$ is $n_1 v_1 + \frac{d}{dx}(n_1 v_1)dx$. The number which disappear by recombination in the layer is $\alpha n_1 n_2 dx$, so that if q is the number of molecules ionized per unit volume per unit time, which we shall suppose constant from one electrode to the other, then

$$\frac{d}{dx}(n_1 v_1) = q - \alpha n_1 n_2.$$

In the same way, for the electrons,

$$-\frac{d}{dx}(n_2 v_2) = q - \alpha n_1 n_2.$$

In the uniform gradient

$$\frac{d}{dx}(n_1 v_1) = \frac{d}{dx}(n_2 v_2) = 0,$$

so that $q = \alpha n^2$, or the ionization is equal to the recombination.

If no ions are emitted by the positive electrode, then $n_1 = 0$ at its surface, that is at $x = 0$. Thus in the layer at the surface of the positive electrode in which X is greater than X_0, the flow of positive ions increases from zero at $x = 0$ to nk_1X_0, its value in the uniform gradient X_0. Let the thickness of this layer be λ_1; then we have

$$\int_0^{\lambda_1} \frac{d}{dx}(n_1 v_1)dx = nk_1 X_0 = \int_0^{\lambda_1} (q - \alpha n_1 n_2)dx.$$

This equation merely expresses the fact that in the layer λ_1 at the positive electrode there is an excess of ionization over recombination sufficient to supply the positive ions which flow across the uniform gradient. In the same way for the

layer near the negative electrode, if this electrode emits no electrons, $n_2 = 0$ at its surface, and

$$-\int_{d-\lambda_2}^{d} \frac{d}{dx}(n_2 v_2)\,dx = nk_2X_0 = \int_{d-\lambda_2}^{d} (q - \alpha n_1 n_2)\,dx,$$

where d is the distance between the electrodes, and λ_2 the thickness of the layer at the negative electrode, in which $q > \alpha n_1 n_2$.

Now $i = e(n_1v_1 + n_2v_2)$, so that $n_1v_1 + n_2v_2$ is constant from one electrode to the other. Thus n_1v_1e increases from 0 at $x = 0$ to enk_1X_0 at $x = \lambda_1$, remains constant from $x = \lambda_1$ to $d - \lambda_2$, and increases from enk_1X_0 at $x = d - \lambda_2$ to i at $x = d$, while n_2v_2e is equal to i at $x = 0$, decreases to enk_2X_0 at $x = \lambda_1$, remains constant from $x = \lambda_1$ to $x = d - \lambda_2$, and decreases from enk_2X_0 at $x = d - \lambda_2$ to zero at $x = d$. Dividing nk_2X_0 by nk_1X_0 we get

$$\frac{k_2}{k_1} = \frac{\int_{d-\lambda_2}^{d} (q - \alpha n_1 n_2)\,dx}{\int_{0}^{\lambda_1} (q - \alpha n_1 n_2)\,dx}.$$

Now $q - \alpha n_1 n_2$ is equal to q at $x = 0$ and to zero at $x = \lambda_1$, and it is equal to zero at $x = d - \lambda_2$ and to q at $x = d$, so that the average value of $q - \alpha n_1 n_2$ over λ_1 cannot differ much from that over λ_2. Hence the ratio of the two integrals must be nearly equal to λ_2/λ_1, so that we get $k_2/k_1 = \lambda_2/\lambda_1$ approximately.

When both electrodes are red hot it is found that λ_2 is much greater than λ_1, so that k_2 must be greater than k_1. λ_1 is so small that it cannot be determined accurately, whereas λ_2 is several millimetres. We should expect the electrons to have a much greater velocity than positive ions, since the mass of an electron is several thousand times smaller than that of an ion.

The drops of potential in the layers λ_1 and λ_2 can be calculated approximately as follows:

The equation $$\frac{dX}{dx} = 4\pi e(n_1 - n_2),$$

multiplied by $$X = \frac{v_1}{k_1} = \frac{v_2}{k_2},$$

gives $$\frac{dX^2}{dx} = 8\pi e\left(\frac{n_1v_1}{k_1} - \frac{n_2v_2}{k_2}\right).$$

Differentiating this with respect to x, and substituting for $\frac{d}{dx}(n_1v_1)$ and $-\frac{d}{dx}(n_2v_2)$ the value $q - \alpha n_1 n_2$, we get

$$\frac{d^2X^2}{dx^2} = 8\pi e\left(\frac{1}{k_1} + \frac{1}{k_2}\right)(q - \alpha n_1 n_2).$$

Now at $x = 0$ we have $n_1 = 0$, and at $x = \lambda_1$, $q = \alpha n_1 n_2 = \alpha n^2$, so that $q - \alpha n_1 n_2$ changes from q at $x = 0$ to 0 at $x = \lambda_1$. For an approximate calculation we may therefore take $q - \alpha n_1 n_2 = q\left(1 - \frac{x}{\lambda_1}\right)$ for values of x between 0 and λ_1. Assuming this, we get

$$\frac{d^2X^2}{dx^2} = 8\pi e\left(\frac{1}{k_1} + \frac{1}{k_2}\right)q\left(1 - \frac{x}{\lambda_1}\right),$$

and by integration
$$\frac{dX^2}{dx} = 8\pi e\left(\frac{1}{k_1} + \frac{1}{k_2}\right)q\left(x - \frac{x^2}{2\lambda_1}\right) + \text{constant}.$$

The equation $\frac{dX}{dx} = 4\pi e(n_1 - n_2)$ at $x = 0$ becomes $\frac{dX}{dx} = -4\pi e n_2$, and also at $x = 0$ we have $i = n_2 e k_2 X$, so that

$$X\frac{dX}{dx}_{x=0} = -\frac{4\pi i}{k_2},$$

or
$$\frac{dX^2}{dx}_{x=0} = -\frac{8\pi i}{k_2}.$$

Using this to determine the constant we get

$$\frac{dX^2}{dx} = 8\pi e\left(\frac{1}{k_1} + \frac{1}{k_2}\right)q\left(x - \frac{x^2}{2\lambda_1}\right) - \frac{8\pi i}{k_2}.$$

At $x = \lambda_1$, X becomes constant so that $\frac{dX^2}{dx} = 0$, therefore

$$\lambda_1 = \frac{2i}{qe}\frac{k_1}{k_1 + k_2},$$

so that
$$\frac{dX^2}{dx} = -\frac{8\pi i}{k_2}\left(1 - \frac{x}{\lambda_1}\right)^2.$$

Integrating this and putting $X^2 = X_0^2$ at $x = \lambda_1$

we get
$$X^2 = \frac{8\pi i\lambda_1}{3k_2}\left(1 - \frac{x}{\lambda_1}\right)^3 + X_0^2.$$

Since X is large compared with X_0 except when x is nearly equal to λ_1, we have approximately from this:

$$X = \sqrt{\frac{8\pi i\lambda_1}{3k_2}\left(1 - \frac{x}{\lambda_1}\right)^3}.$$

Hence the drop of potential in the layer λ_1 is given by

$$V_1 = \int_0^{\lambda_1}\left(\frac{8\pi i\lambda_1}{3k_2}\right)^{\frac{1}{2}}\left(1 - \frac{x}{\lambda_1}\right)^{3/2}dx.$$

This gives
$$V_1 = \left(\frac{8\pi i}{3k_2}\right)^{\frac{1}{2}}\frac{2}{5}\lambda_1^{3/2}.$$

Now
$$\lambda_1 = \frac{2ik_1}{qe(k_1 + k_2)},$$

so that finally
$$V_1 = \frac{16}{5}\left(\frac{\pi}{3k_2}\right)^{1/2}\left(\frac{k_1}{qe(k_1 + k_2)}\right)^{3/2}i^2.$$

In the same way we obtain for the potential drop V_2 at the negative electrode

$$V_2 = \frac{16}{5}\left(\frac{\pi}{3k_1}\right)^{1/2}\left(\frac{k_2}{qe(k_1 + k_2)}\right)^{3/2}i^2.$$

Hence approximately

$$\frac{V_2}{V_1} = \left(\frac{k_2}{k_1}\right)^{1/2}\left(\frac{k_2}{k_1}\right)^{3/2} = \left(\frac{k_2}{k_1}\right)^2,$$

The difference of potential between the electrodes V is therefore given by

$$V = \frac{id}{ne(k_1 + k_2)} + \frac{1}{6}\left(\frac{\pi}{3k_1k_2}\right)^{1/2}\left(\frac{1}{qe(k_1 + k_2)}\right)^{3/2}(k_1^2 + k_2^2)i^2.$$

This equation agrees with that found experimentally, viz:

$$V = ACd + BC^2.$$

The ratio V_2/V_1 cannot be obtained accurately from the observed distribution of the potential between the electrodes when both electrodes are red hot, because V_1 is too small to measure accurately, but it is clear that the ratio is quite large. The electrodes cool the flame so that the ionization close to the electrodes must be less than elsewhere in the flame, whereas we have supposed q constant over the distance between the electrodes.

This approximate theory of the relation between the current and the potential in the flame agrees in a general way with the facts, so that we may say that the results obtained are in accordance with the ionic theory. The two layers λ_1 and λ_2 have a total thickness

$$\lambda_1 + \lambda_2 = \frac{2i}{qe},$$

and we have supposed that there is a uniform potential gradient between the layers. As the potential difference increases i increases, so that $\lambda_1 + \lambda_2$ also increases and so would eventually become equal to d, the distance between the electrodes. The current would then be equal to $\frac{1}{2}qed$, which is one-half the saturation current qed which would be obtained if there were no recombination. The theory therefore is only applicable when the current density is less than one-half the saturation current density.

When the potential difference between red-hot electrodes in a uniform flame is gradually increased, then at first the current is given approximately by the equation

$$V = ACa + BC^2,$$

but, when V becomes equal to a certain large value which depends on the distance between the electrodes, the current begins to increase much more rapidly with V than the above equation indicates. If V is increased much farther an arc discharge starts and the electrodes are melted. This rapid increase of the current is due to ionization by collisions. Ionization by collisions in gases at ordinary temperatures is discussed in the chapter on the motion of electrons in gases, and it obeys the same laws in flame and so need not be further discussed here. The currents obtained before ionization by collisions sets in are probably very small compared with the saturation current.

4. Electron Mobilities in Flames.

We have supposed that the velocity v_2 of the electrons along the direction of the electric field X is given by $v_2 = k_2X$. According to the theory of the motion of electrons in gases we have approximately $k_2 = e\lambda/mV$, where λ is the mean free path of the electrons, e the charge and m the mass of an electron, and V the average velocity of agitation

of the electrons. Townsend's experiments show that in gases at ordinary temperatures V is a function of X/p, where p is the gas pressure, and that the average kinetic energy of agitation of the electrons $\frac{1}{2}mV^2$ is much greater than that of a gas molecule at the same temperature, provided X/p is not extremely small. For example, in nitrogen with X/p equal to 0·25, X being in volts per centimetre and p in millimetres of mercury, he found V to be 7·5 times greater than the value corresponding to the average energy of a gas molecule.

In a flame at about 2000° K. the gas density is about $\frac{1}{7}$ of that at the ordinary temperature, and so is the same as for a gas at the ordinary temperature at about 100 mm. pressure. Thus in a flame with an electric field of 25 volts per centimetre we should expect the velocity of the electrons in the direction of the field to be about the same as Townsend found in nitrogen with $X/p = 0·25$, which was 5×10^5 cm. per second. This makes $k_2 = 5 \times 10^5/25 = 2 \times 10^4$ cm./sec. for 1 volt per centimetre. This estimate is probably too high because the kinetic energy of the electrons in the flame with $X = 0$ is seven times that at the ordinary temperature. If we suppose that the field increases the kinetic energy in the same ratio at any temperature, then with $X = 25$ volts/cm. we get

$$k_2 = 2 \times 10^4/\sqrt{7} = 7500 \,\frac{\text{cm.}}{\text{sec.}} \text{ per } \frac{\text{volt}}{\text{cm.}}.$$

Various attempts to estimate k_2 in flames, by finding the field required to make the electrons move down the flame against the upward stream of gases, and by other similar methods, have been made and results varying from 1000 $\frac{\text{cm.}}{\text{sec.}}$ to 30,000 $\frac{\text{cm.}}{\text{sec.}}$ for 1 volt per centimetre obtained. Such methods are very difficult in flames, because the gases are strongly ionized throughout the flame so that it is difficult to prove that electrons are moving down the flame. Estimates of k_2 by indirect methods are therefore more reliable.

Probably the best method of getting the mobility of the electrons in a Bunsen flame is by measuring the Hall Effect. A flame like that described at the beginning of this chapter is placed between the poles of a large electromagnet, which when excited produces a horizontal field perpendicular to the plane of the flame. Two fine platinum wires are mounted on the end of a shaft which turns in a hole bored in one of the poles of the magnet parallel to the direction of the magnetic field. The two wires are parallel to the magnetic field and about 1 cm. apart, the axis of rotation of the shaft being half-way between them. The ends of the wires project into the flame, and they are insulated and connected to an insulated quadrant electrometer. When a current is passed horizontally between electrodes in the flame so that the

two wires are in the uniform potential gradient X_0 between the electrodes, the quadrant electrometer will indicate a potential difference equal to $X_0 l \sin\theta$, where l is the distance between the two wires, and θ is the angle between the plane containing the two wires and a plane perpendicular to the electric field X_0. By turning the shaft the angle θ can be varied so that it is easy to make the potential difference zero, in which case $\theta = 0$. If now the magnet is excited the quadrant electrometer is deflected, but by rotating the shaft the deflection can be brought back to zero. The magnetic field rotates the equipotential planes in the flame through an angle ϕ which can be measured in this way.

If v_2 is the velocity of the electrons due to the electric field X_0, then the magnetic field will give rise to a vertical force on them equal to Hev_2. Since the flame is insulated, this force cannot produce a vertical current, so that a vertical electric field is produced of strength Z such that

$$Ze = Hev_2.$$

The equipotential planes are therefore turned through the angle ϕ given by

$$\tan\phi = \frac{Z}{X_0},$$

so that

$$v_2 = \frac{Z}{H} = \frac{X_0}{H}\tan\phi.$$

If we put $v_2 = k_2 X_0$, this becomes

$$k_2 = \frac{\tan\phi}{H}.$$

In this way it has been found that k_2 is about $2600 \frac{\text{cm.}}{\text{sec.}}$ for 1 volt per centimetre. This high mobility of the negative ions in flames shows that they must be electrons.

The conductivity of the flame as measured by the ratio of the current to the uniform field X_0 is found to be somewhat diminished by a transverse magnetic field. When the current is horizontal, and the magnetic field is also horizontal and perpendicular to the current, there is a vertical mechanical force on the flame, as with any conductor carrying a current in a magnetic field. This force retards the upward motion of the flame when the magnetic field is in one direction and accelerates it when in the opposite direction.

5. Conductivity of Metallic Vapours in Flames.

When the vapours of certain metallic salts are introduced into a flame the electrical conductivity is greatly increased. The alkali metals of larger atomic weight produce the greatest conductivity.

It is found that the relation between the current, the potential and the distance between the electrodes, and the distribution of potential between the electrodes, for a flame made strongly conducting by an alkali salt vapour are quite similar to those for a flame free from salt. The salt increases the ionization and so the current, but does not alter the nature of the phenomena. The Hall Effect also has about the same value as in a flame free from salt. It appears that the negative ions in a salted flame are electrons, since their velocity due to a field of 1 volt per centimetre is several thousand centimetres per second, as in an unsalted flame.

Interesting effects are produced by putting a bead of salt on a platinum wire into different parts of an otherwise unsalted flame. If the bead is put in anywhere, so that the salt vapour does not come in contact with the negative electrode, then there is no appreciable effect on the current. If, however, the salt vapour is allowed to get to the negative electrode then a very large increase of the current is produced. As we have seen, most of the fall of potential between the electrodes occurs close to the negative electrode, so that we may say that nearly all the electrical resistance of the flame is in the layer close to the negative electrode. The salt vapour therefore has very little effect unless it gets into this layer at the negative electrode.

The potential drop at the negative electrode can be made small or zero by coating the electrode with lime or barium oxide. This causes the electrode to emit electrons, so that if the electrons emitted are sufficiently numerous to carry the current the potential drop disappears. The uniform gradient X_0 then extends right up to the negative electrode, and the relation between the current C and potential difference V becomes roughly

$$V = ACd,$$

instead of $$V = ACd + BC^2.$$

The gradient X_0 is then nearly equal to V/d, which is much greater than before, so that the current is proportionally increased. A similar effect can be produced by putting any alkali metal salt on the negative electrode.

If the potential difference is reversed so that the coated electrode is positive, the current is then given by $V = ACd + BC^2$, and is the same as when both electrodes are of clean platinum. Thus with one electrode coated with lime the flame acts as a rectifier for an alternating current. When the negative electrode is coated with lime the equation $V = ACd$ holds approximately, provided the current C is not greater than the current the electrons emitted by the lime can carry. If V is increased so that C becomes greater than this, the negative drop reappears.

When there is little or no negative drop, putting a bead of salt into

any part of the flame between the electrodes increases the current because the resistance is not then concentrated at the negative electrode. Such experiments show clearly that the salt vapour is strongly ionized in any part of the flame

Different salts of the same alkali metal, for example KCl, K_2CO_3, and KNO_3 give nearly equal conductivities to a flame It is therefore probable that the salts are dissociated and that the metal is present in the flame as metallic vapour. The extreme smallness of the partial pressure of the salt in the flame, which is usually of the order of 10^{-6} mm., makes such a dissociation very probable.

6. Thermodynamical Theory.

The metallic atoms in the flame are ionized by collisions with other atoms, or by the action of light radiation, so that we get electrons and positively charged metallic atoms The electrons and ions recombine, a state of equilibrium being established when the ionization is equal to the recombination. The condition of equilibrium can be obtained by means of the thermodynamical theory of chemical equilibrium in a mixture of gases. It is shown in the chapter on the Quantum Theory (section 7) that the entropy Φ of one mol of a monatomic gas is given by the equation

$$\Phi = k\mathfrak{N} \log \left\{ \frac{V\varepsilon^{5/2}}{\mathfrak{N} h^3} (2\pi m k\theta)^{3/2} \right\},$$

where k is the gas constant for one molecule, $\mathfrak{N}$ the number of molecules of any gas in one mol, V the volume of one mol of the gas, ε the base of Napierian logs, h Planck's constant, θ the absolute temperature, and m the mass of one molecule

Consider a very large quantity of the metallic vapour and let p_1 be the partial pressure of the metallic atoms in it, p_2 the partial pressure of the positive ions, and p_3 that of the electrons. Suppose that the mixture is in a state of equilibrium, and that one mol of the atoms dissociates into ions and electrons. Then, provided the amount of the vapour is so large that this dissociation produces no appreciable change in the partial pressure, the resulting total change of entropy must be zero, by the second law of thermodynamics. Hence

$$\Phi_3 + \Phi_2 - \Phi_1 - \frac{H}{\theta} = 0,$$

where Φ_1 is the entropy of one mol of the atoms, Φ_2 that of one mol of the ions, Φ_3 that of one mol of the electrons, and H the amount of heat energy which must be added to the vapour to keep its temperature constant during the dissociation. For $\Phi_3 + \Phi_2 - \Phi_1$ is the increase in the entropy of the vapour, and H/θ is the entropy lost by the surrounding bodies which supply the heat H. Now, in the metallic vapour, the ions and the electrons are all monatomic gases, so that we may substitute for Φ_1, Φ_2, and Φ_3 the above expression for Φ.

Putting $pV = k\mathfrak{N}\theta$, it becomes

$$\Phi = k\mathfrak{N} \log \left\{ \frac{k\theta\varepsilon^{5/2}}{ph^3} (2\pi m k\theta)^{3/2} \right\},$$

so that we get
$$\Phi_2 - \Phi_1 = k\mathfrak{N} \log \left\{ \frac{p_1}{p_2} \left(\frac{m_2}{m_1}\right)^{3/2} \right\},$$

where m_2 is the mass of one ion and m_1 that of one atom. But m_1 and m_2 are

practically equal, since the mass m_3 of an electron is negligible compared with that of an atom. Hence

$$\Phi_2 - \Phi_1 = k\mathfrak{N}\theta \log \frac{p_1}{p_2}.$$

The equation $\Phi_3 + \Phi_2 - \Phi_1 - \frac{H}{\theta} = 0$ therefore gives

$$\log \frac{p_2 p_3}{p_1} = -\frac{H}{k\mathfrak{N}\theta} + \log \left\{ \frac{(k\theta)^{5/2}}{h^3} (2\pi m_3 k\theta)^{3/2} \right\}.$$

Now H is equal to $H_0 + k\mathfrak{N}\theta$, where H_0 is the increase in the internal energy due to the dissociation, for $k\mathfrak{N}\theta = pV$ is the external work done. Let $H_0 = \mathfrak{N}Pe$, where e is the charge on one electron, and P the potential difference through which a charge e must fall to acquire enough energy to dissociate one of the metallic atoms into an ion and an electron.

Then, putting $K = \frac{p_2 p_3}{p_1}$, we get

$$\log K = -\frac{Pe}{k\theta} + \tfrac{5}{2} \log \theta + \log \left\{ \frac{k^{5/2}}{h^3} (2\pi m_3)^{3/2} \right\}.$$

All the quantities in the last term on the right-hand side of this equation are known, so that we can calculate K at any temperature θ, provided we know P for the metal vapour.

7. Conductivity with Varying Amounts of Salt in Flame.

In the flame there is some ionization of the flame gases in addition to the ionization of the metal vapour. Let p_2' be the partial pressure of the positive ions formed from the flame gases, and p_1' the partial pressure of the undissociated flame molecules, and let $K' = \frac{p_2' p_3}{p_1'}$ be the equilibrium constant for the flame molecules.

The conductivity of the flame is proportional to the number of electrons, so that we have $p_3 = Ac$, where A is a constant and c denotes the conductivity. Also $p_3 = p_2 + p_2'$, since the number of electrons must be equal to the total number of ions. Let $p = p_1 + p_2$ be the partial pressure of the metal atoms, neutral and dissociated, also let $p' = p_1' + p_2'$.

We shall assume p_2' to be very small compared with p_1', so that approximately $p' = p_1'$.

Now $\quad K = \frac{p_2 p_3}{p - p_2}, \quad$ or $\quad p_2 = \frac{Kp}{K + Ac}.$

Also $\quad K' = \frac{p_2' p_3}{p'}, \quad$ so that $\quad p_2' = \frac{K'p'}{Ac}.$

Hence the equation $\quad Ac = p_2 + p_2'$

becomes $\quad Ac = \frac{Kp}{K + Ac} + \frac{K'p'}{Ac}.$

When there is no metal vapour in the flame so that $p = 0$, let $c = c_0$, so that

$$Ac_0 = \frac{K'p'}{Ac_0};$$

then putting $\frac{c}{c_0} = x$, we get

$$A\left(x - \frac{1}{x}\right) = \frac{Kp}{c_0(K + Ac_0x)},$$

or

$$\frac{px}{x^2 - 1} = Ac_0\left(1 + \frac{Ac_0x}{K}\right).$$

The partial pressure p is proportional to the amount of salt in the flame, so that this equation can be tested by observing how the conductivity varies with the amount of salt in the flame. When x is large it becomes approximately $Kp = (Ac_0x)^2$, and the conductivity should therefore be proportional to the square root of the amount of salt present. A definite amount of any salt can be introduced into a flame by spraying a solution of the salt by means of a jet of compressed air, mixing the air and spray with coal gas and then passing it into the burner. In this way the salt is uniformly distributed through the flame, and the amount of it entering the flame is proportional to the concentration of the solution used. By finding the amount of solution used up in a known time the amount of salt entering the flame in unit time can be determined.

If g is the weight of salt per unit volume in the salt solution sprayed into the flame we may put $p = Bg$, where B is a constant, so that

$$\frac{gx}{x^2 - 1} = \frac{Ac_0}{B}\left(1 + \frac{Ac_0x}{K}\right).$$

If $x = c/c_0$ is found for a series of values of g, then on plotting $gx/(x^2 - 1)$ against x a straight line given by $y = b + ax$ should be obtained, where $y = \frac{gx}{x^2 - 1}$, $b = \frac{Ac_0}{B}$, and $a = \frac{A^2c_0^2}{BK}$. Hence $K = B\frac{b^2}{a}$. The conductivity c of the flame is proportional to the ratio of the current to the uniform potential gradient between the electrodes, so that $x = c/c_0$ can be easily determined.

The following table gives the results of a series of measurements of x for a flame into which solutions of caesium chloride were sprayed.

Grammes CsCl per Litre in Solution (g).	x.	$\frac{10^4 gx}{x^2 - 1}$	$10 + x$
0	1	—	—
0·0032	2·88	12·6	12·88
0·008	5·72	14·5	15·72
0·016	8·9	18·2	18·9
0·032	13·5	23·9	23·5
0·08	22·7	36·0	32·7
0·16	32·8	49·0	42·8
0·8	85·2	94·0	95·2
8·0	282·0	284·0	292·0
80·0	883·0	906·0	893·0

It appears that $\frac{10^4gx}{x^2 - 1}$ is nearly equal to $10 + x$, so that the conductivity

of the flame varies with the amount of metal vapour in it, approximately in accordance with the ionic theory. Similar results have been obtained with other alkali metals.

The equation $p_2 = \frac{Kp}{K + Ac}$ gives for the fraction of the metal atoms ionized $\frac{p_2}{p} = \frac{K}{K + Ac}$, which is equal to $\frac{b}{b + ax}$. In the case of caesium therefore in the particular flame used the fraction of the caesium atoms ionized was $\frac{10}{10 + x}$. With an indefinitely small amount of caesium $x = 1$, so that the fraction ionized is 10/11. The ionization does not become complete because there are some electrons in the flame when no caesium is present.

If the number of metal atoms in unit volume of the flame is n and the corresponding conductivity c, the number of electrons in unit volume is $n_3 = nbx/g$, for we have $p_3 = Ac_0x$ and $p = Bg$, and therefore

$$\frac{p_3}{p} = \frac{n_3}{n} = \frac{Ac_0x}{Bg},$$

and $b = Ac_0/B$. Thus the number n_3 of electrons in unit volume of the flame can be calculated from the ratio n/g. To find n/g it is necessary to find the amount of solution entering the flame, and the volume of the flame gases with which it is mixed. The volume of the flame gases can be found from the horizontal cross-section of the flame and the upward velocity of the flame gases. In this way n/g can be estimated and so n_3 found. The current density in the flame in the uniform potential gradient is $n_3ek_2X_0$, so that when n_3 is known k_2, the mobility of the electrons, can be obtained. In this way it has been found that k_2 is about 2600 cm. per second for 1 volt per centimetre.

The equilibrium constant K can be calculated by means of $K = Bb^2/a$. We have $p = Bg = nk\theta$, where n is the number of metal atoms per unit volume in the flame, k the gas constant for one molecule, and θ the absolute temperature. This gives B when n is known, and K can then be calculated. It is found that the values of K found in this way for the alkali metals are nearly equal to those given by the thermodynamical theory of the equilibrium between the atoms, ions, and electrons. This shows that the quantum theory of the entropy of electron gas is approximately correct.

8. Conductivity for Alternating Currents.

The conductivity of flames for rapidly alternating currents has been investigated with interesting results. Suppose we have two plane parallel electrodes in a Bunsen flame at a distance d apart, and that an alternating potential difference is maintained between them. Let the electrodes be symmetrically placed in the flame and red hot.

The mobility of the positive ions in the flame is very small compared with that of the electrons, and their mass is enormously greater. Clearly, then, in a rapidly alternating electric field the amplitude of vibration of the ions will be very small compared with that of the electrons. We shall therefore assume that the positive ions do not move at all and that the current is all carried by the electrons. When there is no electric field, let there be n positive ions per unit volume throughout the space between the electrodes, and an equal number of

electrons. When the alternating field is applied let the electrons vibrate with an amplitude A. All the electrons within a distance A of the electrodes will therefore strike the electrodes and be removed from the flame. We suppose $2A$ less than d, so that a layer of electrons of thickness $d - 2A$ remains in the flame and oscillates between the electrodes.

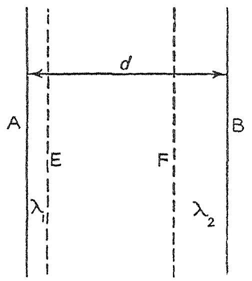


FIG 2

In the layer of electrons the density of charge is zero, but outside the layer it is ne where e is the charge on one positive ion.

In fig. 2 A and B are the electrodes, and the layer of electrons is between E and F. Let the layer near A in which there are no electrons be of thickness λ_1, and that near B of thickness λ_2. Then $\lambda_1 + \lambda_2 = 2A$, and if y denotes the displacement of the layer of electrons from its mean position then

$$\lambda_1 - \lambda_2 = 2y.$$

Let V be the potential difference between the electrode A and a point at a distance x from it. From $x = 0$ to $x = \lambda_1$ and from $x = d - \lambda_2$ to $x = d$, we have $\frac{\partial^2 V}{\partial x^2} = -4\pi\rho$ where $\rho = ne$. Also from $x = \lambda_1$ to $x = d - \lambda_2$ we have $\frac{\partial^2 V}{\partial x^2} = 0$. By integrating these equations it is easy to calculate V at $x = d$. The integration constants are found by making V and $\frac{\partial V}{\partial x}$ continuous at $x = \lambda_1$ and $x = d - \lambda_2$. If X denotes the field strength in the space between E and F, we find for the potential difference between the electrodes

$$V = -8\pi\rho Ay - Xd.$$

The term Xd will probably be small unless ρ is very small, so that approximately when ρ is large $V = 8\pi\rho Ay$.

The equation of motion of the electrons is

$$m\frac{d^2y}{dt^2} = -Xe - \mu\frac{dy}{dt},$$

where m is the mass and $-e$ the charge of one electron, and μ is the average viscous resistance to the motion for unit velocity. We shall suppose that μ is small, so that approximately

$$m\frac{d^2y}{dt^2} = -Xe.$$

If then $X = X_0\epsilon^{ipt}$, and $y = A\epsilon^{ipt}$, then $A = -eX_0/mp^2$. Putting $V = V_0\epsilon^{ipt}$, and substituting in $V = 8\pi\rho Ay$, we get

$$A = \sqrt{\frac{V_0}{8\pi\rho}}.$$

The current density I is given approximately by

$$I = -\rho\frac{dy}{dt} = -\rho\sqrt{\frac{V_0}{8\pi\rho}}\,ip\epsilon^{ipt}.$$

This shows that the electrodes in the flame behave like a condenser. For the charge on a condenser of capacity C is $Q = VC = CV_0\epsilon^{ipt}$, and the current charging it is $\frac{dQ}{dt} = CV_0ip\epsilon^{ipt}$. Comparing this with the expression for I we see that the electrodes in the flame have an apparent capacity C per unit area given by

$$C = \sqrt{\frac{\rho}{8\pi V_0}}.$$

The apparent capacity of the electrodes in the flame can be measured by means of a high-frequency Wheatstone bridge in which the capacity of the electrodes is balanced by three air condensers. In this way it has been found that C is nearly inversely as $\sqrt{V_0}$ and directly as $\sqrt{\rho}$, in accordance with the theory. If the viscous resistance to the motion of the electrons is not neglected, then the theory indicates that the electrodes should behave like a condenser shunted by a high resistance. By determining the apparent resistance it is possible to estimate the viscous resistance to the motion of the electrons and so get an estimate of their mobility. The results obtained in this way are of the same order as those given by the other methods.

9. Mobility of Positive Ions in Flames.

The mobility of the positive ions of alkali metals in flames has been estimated by finding the strength of the electric field required to make them move down the flame against the upward stream of gases.

In fig. 3 FF is a Bunsen flame. A and B are two electrodes consisting of fine platinum wire gratings with wires about 0·5 cm. apart which are supported one above the other in the flame. A bead of salt on a wire D can be introduced into the flame just below the upper electrode. The upper electrode is charged positively and the current from it to a wire C, the end of which is in the plane of B at the axis of the flame, is measured with a galvanometer. The wire C is connected to one terminal of the galvanometer and the other terminal is connected to the electrode B, which is connected to the earth. The current is measured with and without the salt bead in the flame. It is found

that when the potential difference between A and B exceeds a certain value, the current down the axis of the flame to C is increased by putting in the bead. This is supposed to indicate the point at which the ions from the salt move down When care is taken to prevent salt vapour from the bead getting into the lower parts of the flame, it is found that the electric field required to make the ions move down is about 200 volts per centimetre. The upward velocity of the flame gases is about 200 to 300 cm. per second, so that the mobility of the ions is about 1 cm. per second for 1 volt per centimetre The mobility of the electrons being about 2600 cm per second for 1 volt per centimetre, it follows that practically all the current is carried by the electrons, except close to the negative electrode It is found that the mobility of the positive ions is practically the same for lithium, sodium, potassium, rubidium, and caesium ions.

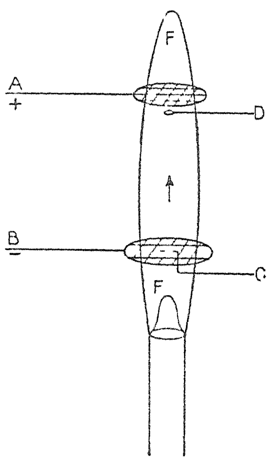


Fig 3

The mobility of an electron is nearly equal to $e\lambda/mV$, where e is the charge and m the mass of an electron, λ the mean free path, and V the average velocity of agitation. This expression was obtained by assuming that after a collision of an electron with a molecule the velocity of the electron was equally likely to be in any direction. This assumption cannot be made in the case of an ion, because the mass of an ion is comparable with that of a molecule.

If the velocity of an ion is given by $v = kX$, where X is the field, the force on an ion moving with unit velocity through the gas must be $Xe/v = e/k$. If an ion of mass M is projected into the gas with initial velocity V_0, the average distance it will travel in its original direction may be calculated as follows.

Let a large number N of ions be projected into the gas with average initial velocity V_0. Then

$$\frac{d}{dt}(NMV) = -\frac{e}{k}NV,$$

where V is the average velocity of the N ions in the original direction. This gives

$$V = V_0 \epsilon^{-\frac{e}{kM}t}.$$

The average distance L which they go in the original direction is

$$L = \int_0^\infty V\,dt = V_0 \int_0^\infty \epsilon^{-et/kM}\,dt = \frac{V_0 kM}{e}.$$

Thus $k = \frac{eL}{MV_0}$, so that if V_0 is equal to the average velocity of agitation of the ions, which we denote by V, then

$$k = \frac{eL}{MV}.$$

In the case of an electron $L = \lambda$, since the electrons lose all their original velocity on the average at the first collision The positive ions of the different alkali metals in a flame all have about equal mobilities, so that it appears that L is proportional to MV. or to $\sqrt{M}$, since V is inversely as $\sqrt{M}$. The mobility of the positive ions in flames seems to be about the same as that of the positive ions produced by X-rays in air at atmospheric pressure and at the ordinary temperature.

REFERENCES

1. *Electrical Properties of Flames* H A. Wilson.

2. *Handbuch der Experimentalphysik* Wien und Harms Band 13, 1 Teil. "Die Elektrischen Eigenschaften der Flamme." A Becker

3. "Electricity in Flames" J A J Bennett, *Phil Mag*, January, 1927.

4. "The Electrical Conductivity of Flames." H A. Wilson. *Physical Review Supplement*, 1930.

CHAPTER XVI

The Positive Column and Negative Glow

1. The Positive Column, Negative Glow, Crookes Dark Space, and Faraday Dark Space.

When an electric discharge is passed between electrodes in a gas at a pressure of the order of 1 mm. of mercury, a luminous column extending some distance from the positive electrode or anode is usually obtained which is called the positive column. Near the negative electrode or cathode there is a luminous layer separated from the cathode by a non-luminous region which is called the Crookes dark space. The luminous layer is sharply defined on the side near the cathode, but fades gradually on the other side with increasing distance from the cathode. This luminous layer is called the negative glow.

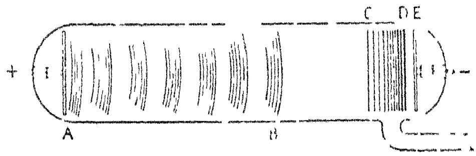


Fig 1

Between the negative glow and the positive column there is a non-luminous region called the Faraday dark space. The colour of the light emitted by the negative glow is often quite different from that emitted by the positive column.

When the distance between the electrodes is increased, keeping the gas pressure and the current constant, the distances of the negative glow and of the end of the positive column from the cathode remain unchanged, so that the length of the positive column increases by an amount equal to the increase in the distance between the electrodes.

The luminosity of the positive column is sometimes uniform along its whole length, but frequently varies periodically so that it consists of a series of equally spaced bright layers or striæ separated by more or less dark intervals. Large currents and low pressures favour the appearance of striæ. Fig. 1 shows such a discharge in nitrogen gas. The discharge is produced in a glass tube about 2 cm. in diameter and 20 cm. long, having aluminium disc electrodes sealed in at each end as shown. The positive column, consisting of 8 striæ, extends from A to

B. The Faraday dark space is between B and C, the negative glow between C and D, and the Crookes dark space between D and E. The pressure is about ½ mm. and the current about 5 milliamperes in such a discharge as this.

As the pressure in such a discharge tube is reduced, the positive column gets shorter and finally disappears except for a layer of luminosity on the anode. The negative glow and Crookes dark space get longer, and eventually the Crookes dark space extends to the anode when the pressure is about 0·003 mm. The potential difference required to maintain the discharge depends on the gas pressure. With a discharge tube 3 cm. in diameter having electrodes 11·5 cm. apart and a constant current of 10 milliamperes Townsend obtained the following results:

Pressure in millimetres of mercury	4·0	2·84	1·65	1·04	0·66	0·4	0·29	0·24	0·13
Potential difference in volts	650	620	500	470	490	530	590	630	800

As the pressure is reduced the potential difference falls to a minimum and then rises again. At very low pressures the potential difference required to produce a discharge becomes very large, and it is possible to obtain so good a vacuum that two hundred thousand volts will not produce any discharge.

2. Potential Differences in the Tube.

The potential at any point in a discharge tube may be estimated by means of a small insulated electrode of fine wire projecting into the tube. The wire is usually enclosed in a glass tube except near its end. The potential difference between the wire and one of the electrodes can be measured with an electrostatic voltmeter. Such an insulated

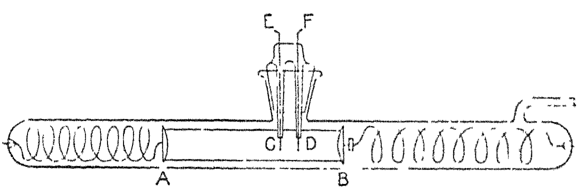


Fig. 2

wire takes up a definite potential which is probably not much different from the potential of the discharge close to it. The current in the discharge is carried by electrons moving towards the anode and positive ions moving towards the cathode. The electrons have much higher velocities than the ions. The wire electrode takes up a potential such that the negative charge it receives by absorbing electrons is equal to the positive charge it receives from the positive ions. It therefore probably takes up a potential somewhat below that of the gas near it

so that it repels electrons and attracts ions, for owing to the higher velocity of the electrons more electrons move towards the wire than ions.

The strength of the electric field along the discharge can be found by means of two similar small electrodes a few millimetres apart. The potential difference between them can be measured with an insulated quadrant electrometer, and the field strength is equal to the potential difference divided by the distance between the electrodes.

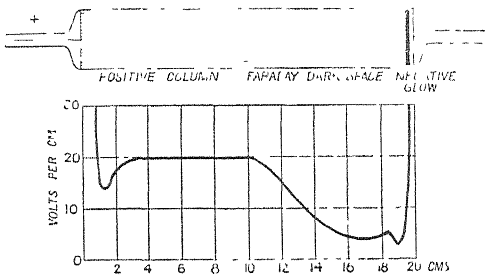


Fig. 3

A discharge tube used by the writer for such measurements is shown in fig. 2. A and B are two aluminium disc electrodes kept at a constant distance apart by three thin glass rods, and connected to wires sealed through the ends of the tube by spiral springs of fine wire. A small piece of iron enables the electrodes to be moved along the tube by means of a magnet. C and D are two fine wire electrodes enclosed in small glass tubes except close to their ends and sealed into a glass stopper as shown. By turning the stopper the distance between the electrodes measured along the axis of the tube can be varied. The electrodes C and D were connected to an insulated quadrant electrometer, and a discharge was passed between A and B from a battery of a large number of small cells.

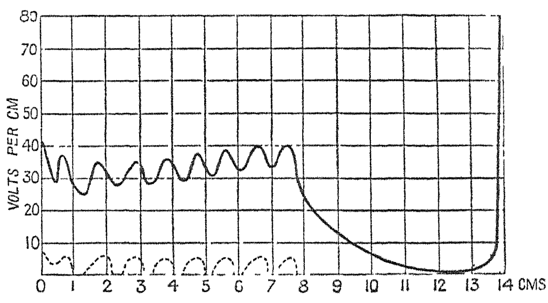


Fig. 4

The distribution of the electric field along the discharge is shown in figs. 3 and 4, which represent results obtained by Graham and by the writer respectively.

In fig. 3 the field strength is constant along the uniform positive column. It rises rapidly near both electrodes and is small in the negative glow and Faraday dark space.

In fig. 4 the dotted curve shows approximately the distribution of luminosity. In this case the field strength rises and falls in the positive

column, the maxima being in the striæ and the minima between them.

In discharges at rather low pressure, about 0·1 mm., when the Faraday dark space extends almost to the anode, the field strength close to the anode becomes very small or even negative.

3. The Cathode Fall of Potential.

There is always a considerable difference of potential between the negative glow and the cathode. This is called the cathode fall of potential. There is also a smaller fall of potential very close to the surface of the anode. The electric field is therefore very large and positive close to the surface of the anode but is very small or even negative 1 or 2 mm from it. The variation of the field near the anode is similar to that near the cathode but is spread over a much smaller space. At the cathode positive ions strike the metal surface and liberate electrons, whereas at the anode electrons strike the metal surface and may liberate some positive ions from it.

The cathode fall of potential can be determined by measuring the potential difference between the cathode and a small wire electrode in the negative glow by means of an electrostatic voltmeter. The negative glow does not cover the whole surface of the cathode when the current is not too large. As the current is increased, the area of the cathode covered by the negative glow is proportional to the current, so that the current density in the glow remains constant. When the glow covers the whole cathode it becomes brighter as the current is increased. The cathode fall of potential is independent of the current and of the gas pressure so long as the cathode is only partially covered by the glow, but increases with the current when the cathode is covered. The cathode fall of potential when the cathode is only partly covered is called the normal cathode fall of potential.

The following table gives some values of the normal cathode fall of potential for cathodes made of different metals.

Gas.	Platinum.	Aluminium.	Potassium.
Oxygen ..	369 volts		
Hydrogen ..	295 ,,	190 volts	172 volts
Nitrogen ..	232 ,,	224 ,,	170 ,,
Helium ..	226 ,,	--	60 ,,
Argon .	167 ,,	-	-

The normal cathode fall of potential is nearly equal to the minimum sparking potential, that is, the sparking potential between parallel plate electrodes when the gas pressure is adjusted so that the sparking potential is as small as possible. The large influence of the nature of

the cathode on the normal cathode fall of potential is probably due to the emission of electrons by the cathode when bombarded by positive ions.

Aston and Watson made a number of measurements of the potential difference required to maintain a discharge between parallel plate electrodes. In their experiments the negative glow extended to the anode so that there was no positive column, and the potential difference between the electrodes was practically equal to the cathode fall of potential. The cathode was entirely covered by the glow, so that the cathode fall of potential was greater than the normal value.

It was found that the potential difference V was given by the following equation

$$V = E + \frac{F\sqrt{C}}{p},$$

where E and F are constants, C is the current density, and p the gas pressure.

Aston and Watson also measured the length D of the Crookes dark space and found that

$$D = \frac{A}{p} + \frac{B}{\sqrt{C}},$$

where A and B are constants. The values of the constants E, F, A, and B depend on the nature of the gas and of the cathode. With D in centimetres, p in $\frac{1}{100}$ mm. of Hg, C in 10^{-4} ampere per square centimetre, the following table gives some values of these constants.

Cathode.	Oxygen.				Hydrogen			
	A.	B.	E.	$F \times 10^{-2}$.	A.	B.	E.	$F \times 10^{-2}$
Aluminium	5 7	0 43	310	17 5	23	0 41	170	66
Copper .	8·9	0·40	340	28·5	47	0 45	300	130
Platinum ..	8·8	0·40	335	30 0	45	0·42	270	120

It appears that, for a given gas and cathode, V and the product pD are functions of $\sqrt{C}/p$. If we eliminate $\sqrt{C}/p$ from the equations giving V and D we get

$$V = E + \frac{FB}{pD - A},$$

so that V depends on pD only in any given case The product pD is proportional to the amount of gas per unit area in the Crookes dark space, and V is nearly equal to the fall of potential across the Crookes

dark space, since the field strength in the negative glow is quite small. It is also found that the sparking potential between parallel plates is a function of the amount of gas per unit area between the plates.

The electric field strength in the Crookes dark space has been investigated by several physicists by observing the potentials taken up by a wire electrode at different distances from the cathode. This method is unreliable when the wire is close to the cathode, because the wire stops the stream of positive ions moving towards the cathode and so prevents the emission of electrons by the cathode immediately behind the wire. The wire therefore becomes positively charged and so does not take up the potential of the gas.

Aston determined the electric field in the Crookes dark space by measuring the deflection of a narrow beam of cathode rays sent across the dark space parallel to the surface of the cathode. In this way he found that the field strength is nearly proportional to the distance from the negative glow, so that $F = Ax$, where F is the field strength, x the distance from the glow, and A a constant. The cathode fall of potential V is therefore given by

$$V = \int_0^D Ax\,dx = \frac{AD^2}{2}.$$

Aston determined A by measuring F at different distances from the negative glow and found that $\frac{1}{2}AD^2$ was equal to the cathode fall of potential determined by measuring the potential difference between the negative glow and the cathode.

4. Theory of the Cathode Fall of Potential.

An approximate theory of the cathode fall of potential can be worked out if we assume that the gas in the Crookes dark space is ionized by collisions with electrons and with positive ions, and that electrons are liberated by collisions of positive ions with the cathode. We also assume that recombination of the ions and electrons can be neglected, and that the number of ions and electrons lost by diffusion to the walls of the discharge tube is inappreciable. The condition which must be satisfied in order that a steady current may be maintained through a gas between parallel electrodes with these assumptions, when no electrons are set free at the cathode, was worked out by Townsend. A slight modification of Townsend's theory gives the condition when electrons are supposed set free at the cathode.

Let n_1 be the number of positive ions per cubic centimetre at a point, and v_1 their velocity; and let n_2 and v_2 be the corresponding quantities for the negative electrons. Let α denote the number of molecules ionized by a negative electron in moving 1 cm., and let β be the corresponding number for a positive ion. The electric intensity X varies with the distance x from the cathode, so that α, β, v_1, and v_2 are not constant. In the steady state we have

$$-\frac{d}{dx}(n_2v_2) + \alpha n_2v_2 + \beta n_1v_1 = 0,$$

and

$$\frac{d}{dx}(n_1v_1) + \alpha n_2v_2 + \beta n_1v_1 = 0.$$

Also, the current density C is equal to $e(n_1v_1 + n_2v_2)$ and is constant. Substituting $C/e - n_1v_1$ for n_2v_2, we get

$$\frac{d}{dx}(n_1v_1) - (\alpha - \beta)n_1v_1 = -\alpha C/e.$$

Let

$$\epsilon^{\int_0^x (\alpha-\beta)dx} = Z,$$

so that

$$\frac{d}{dx}(n_1v_1Z^{-1}) = -\frac{\alpha C Z^{-1}}{e}.$$

Then

$$n_1v_1 = BZ - \frac{C}{e}Z\int_0^x \alpha Z^{-1}dx,$$

where B is constant.

The number of positive ions striking the cathode per square centimetre per second is equal to n_1v_1 at $x = 0$. Let γn_1v_1 be the number of electrons set free by the impacts of these positive ions on the cathode. Then at $x = 0$ we have $n_2v_2 = \gamma n_1v_1$, or

$$n_1v_1 = \frac{C}{(1+\gamma)e}.$$

This condition gives

$$B = \frac{C}{(1+\gamma)e}.$$

Let the distance between the electrodes be S, so that at $x = S$, $n_1v_1 = 0$. This gives

$$0 = \frac{CZ}{e(1+\gamma)} - \frac{CZ}{e}\int_0^S \alpha Z^{-1}dx,$$

or

$$\int_0^S \alpha Z^{-1}dx = \frac{1}{1+\gamma},$$

which is the condition required. If $\gamma = 0$, this reduces to the condition given by Townsend.

In a discharge at moderately low pressure, when a Crookes dark space exists, the electric intensity is very small in the negative glow, so that if the positive electrode is anywhere in the negative glow then

$$\int_0^S \alpha Z^{-1}dx = \int_0^D \alpha Z^{-1}dx,$$

where D is the length of the Crookes dark space, because α and β are both zero when X is very small.

Let $\alpha = \alpha_1 p$, $\beta = \beta_1 p$, $x = y/p$, and $Z_1 = \epsilon^{\int_0^y (\alpha_1-\beta_1)dy}$, so that

$$\int_0^D \alpha Z^{-1}dx = \int_0^{pD} \alpha_1 Z_1^{-1}dy = \frac{1}{1+\gamma}.$$

Also let the field strength $X = pY$. The equation $X = A(D - x)$ becomes $Y = \frac{A}{p^2}(pD - y)$, and, putting $A_1 = A/p^2$, we get $Y = A_1(pD - y)$. The equation $V = \int_0^D X dx$ then becomes

$$V = \int_0^{pD} Y dy = A_1\frac{(pD)^2}{2}.$$

For the normal cathode fall of potential in hydrogen, Skinner found $V = 197$ volts, and $pD = 1.10$. In this case, therefore, $A_1 = \frac{2 \times 197}{(1.10)^2} = 325$, whence $Y = 325(pD - y)$.

Now Townsend has determined α and β as functions of $Y = X/p$, so that by using his values of these quantities we can compute the value of $\int_0^{pD} \alpha_1 Z_1{}^{-1} dy$. It is found in this way that this integral is nearly equal to unity, and γ must therefore be very small.

The equation $n_1 v_1 = BZ - \frac{C}{e} Z \int_0^x \alpha Z^{-1}\, dx$, with $B = \frac{C}{(1+\gamma)e}$, gives

$$n_1 v_1 = \frac{C}{e} Z \left(\frac{1}{1+\gamma} - \int_0^x \alpha Z^{-1}\, dx \right),$$

where x is the distance from the cathode. Also $n_2 v_2 = \frac{C}{e} - n_1 v_1$. Hence, substituting for n_1 and n_2 in the equation

$$\frac{dX}{dx} = 4\pi e (n_1 - n_2),$$

we get, on putting $v_1 = k_1 X$ and $v_2 = k_2 X$,

$$\frac{dX^2}{dx} = -8\pi C \left\{ \frac{1}{k_2} - \left(\frac{1}{k_1} + \frac{1}{k_2} \right) Z \left(\frac{1}{1+\gamma} - \int_0^x \alpha Z^{-1}\, dx \right) \right\}.$$

Let $X = pY$, $px = y$, $k_2 = K_2/p$, $k_1 = K_1/p$, $\alpha = \alpha_1 p$, $\beta = \beta_1 p$, then

$$\frac{dY^2}{dy} = -\frac{8\pi C}{p^2} \left\{ \frac{1}{K_2} - \left(\frac{1}{K_1} + \frac{1}{K_2} \right) Z_1 \left(\frac{1}{1+\gamma} - \int_0^y \alpha_1 Z_1{}^{-1} dy \right) \right\}.$$

where $Z_1 = \varepsilon^{\int_0^y (\alpha_1 - \beta_1)\, dy}$, as before.

Now K_2, K_1, α_1, and β_1 are all functions of Y only, for Townsend has proved experimentally that α/p, β/p, v_1, and v_2 are all functions of X/p. Also γ is presumably a function of X/p only, since it must depend on the velocity of the positive ions at $x = 0$. It appears therefore that the above equation is a relation between the three quantities Y, y, and C/p^2. Hence we may write

$$Y = \phi(y, C/p^2),$$

where ϕ denotes some function of C/p^2 and y only.

Now the cathode fall of potential V is equal to $\int_0^D X\, dx$, and therefore

$$V = \int_0^{pD} Y dy.$$

For a given value of C/p^2, Y is a function of y only, so that pD is a function of C/p^2 only, for pD is the value of y at which Y becomes very small. Hence it follows that V is a function of C/p^2 only. If V has a minimum value when C/p^2 is varied, this minimum value will be independent of p, and the corresponding values of C will be proportional to p^2, and the corresponding values of D will be inversely as p.

5. Comparison with Experiment.

These results deduced from the theory are precisely those which follow from the experiments of Aston and Skinner. Thus Aston found pD and V both to

depend on C/p^2, and Skinner found the current density to be proportional to p^2, and D to be inversely as p for the normal cathode fall, which must be a minimum value of V.

At $y = 0$, the equation for $\frac{dY^2}{dy}$ gives approximately

$$\frac{dY^2}{dy} = -\frac{8\pi C}{K_1 p^2},$$

since K_2 is very large compared with K_1.

If now we assume that $Y = A_1(pD - y)$, we get

$$\frac{dY^2}{dy} = -2A_1^2(pD - y).$$

At $y = 0$, this gives

$$\frac{dY^2}{dy} = -2A_1^2 pD.$$

Also $V = \int_0^{pD} Y dy$, so that $A_1 = 2V/p^2D^2$,

and therefore $$K_1 = \frac{\pi C(pD)^3}{V^2 p^2}.$$

The following table gives the values of pD, C/p^2, and V found by Skinner in hydrogen, and the calculated values of K_1. C is in milliamperes per square centimetre, D in centimetres, and p in millimetres of mercury.

C/p^2.	pD.	V.	K_1.
0·0742	1·100	197	$7·2 \times 10^3$
0·1484	0 844	204	6 1 ,,
0 2968	0 624	227	3 9 ,,
		Mean ..	$5 7 \times 10^3$

The values of K_1 are for 1 volt per centimetre at 1 mm. pressure.

If we assume that the velocity of the positive ions is inversely as the pressure, this gives, for the velocity due to 1 volt per centimetre at 760 mm., 7·5 cm. per second. This is nearly equal to the velocity of positive ions in hydrogen at 760 mm. as found by Zeleny, Langevin, and others.

It appears that the results obtained in the Crookes dark space are consistent with the theory that the ionization is due to collisions, and that recombination and diffusion can be neglected.

The ionization due to the positive ions β is small compared with that due to the electrons, so that if we put $\beta = 0$, the value of the integral $\int_0^{pD} \alpha_1 Z_1^{-1} dy$ is not much affected, but must then be less than unity so that γ may not be zero. It is probable that γ is not exactly zero, and it is difficult to distinguish between ionization by collisions of the ions with the gas molecules, and the liberation of electrons from the cathode by the impact of positive ions. It is quite possible that β is really zero, and that discharges in gases are maintained by the liberation of electrons from the cathode and not by ionization by collisions of positive ions and gas molecules.

6. The Positive Column.

The electric intensity X in the uniform positive column depends on the gas pressure p, the current density C, and the diameter of the discharge tube. In wide tubes, the positive column may not fill the whole cross-section of the tube, and then has a certain cross-section which depends on the pressure and total current. This indicates that as the current density increases the electric intensity diminishes to a minimum value, and then increases as the current density is further increased. The cross-section of the positive column when it does not fill the tube increases with the total current, and the current density in it probably does not vary much as the current is increased.

The electric intensity X is nearly proportional to the square root of the pressure at pressures up to 2 or 3 mm., but at higher pressures is nearly a linear function of the pressure. For example, in air from $p = 0{\cdot}2$ mm. to $p = 2{\cdot}82$ mm., $X = 35\sqrt{p}$, and in hydrogen from $p = 0{\cdot}25$ mm. to $p = 1{\cdot}36$ mm., $X = 28\sqrt{p}$.

When the current density is very small X increases rapidly with the current, but soon attains a nearly constant value as the current is increased. With larger current densities above about 1 milliampere per square centimetre the electric intensity diminishes slightly as the current is increased.

In a uniform positive column the equation $C = e(n_1v_1 + n_2v_2)$ becomes $C = ne(v_1 + v_2)$, because $\frac{dX}{dx} = 4\pi e(n_1 - n_2) = 0$, so that $n_1 = n_2 = n$. The velocity of the electrons v_2 due to the electric field X is much greater than v_1, so that approximately $C = nev_2$. The velocity v_2 has been determined by Townsend, and is a function of $Y = X/p$. Since X is nearly independent of C, except when C is very small, it follows that n must be proportional to C. The rate of production of electrons and ions in a uniform column must be equal to the rate at which they disappear by recombination and diffusion to the walls of the tube. The rate of production may be assumed to be proportional to the electrical energy dissipated per cubic centimetre, or to CX. The electric field does work on the electrons, giving them kinetic energy. When the velocity of an electron is small it loses very little energy by collisions, but when the velocity exceeds a critical value collisions result in ionization or excitation of the molecule with loss of the kinetic energy of the electron. Since C is proportional to n, and X is nearly constant, it follows that the rate of production of ions and electrons, and therefore also the rate of disappearance, must be nearly proportional to n. Recombination is proportional to n^2, so that it appears that the loss by diffusion to the walls of the tube must be large compared with the loss by recombination. We have therefore approximately $\frac{dn}{dt} =$

$ACX - Bn$, where ACX represents the rate of production of ions, and Bn the rate of loss. In the uniform column in a steady state $\frac{dn}{dt} = 0$, so that $X = \frac{Bn}{AC} = \frac{Bn}{Anev_2} = \frac{B}{Aev_2}$. If we take $v_2 = K\frac{X}{p}$, this gives

$$X = \frac{Bp}{AeKX},$$

or

$$X^2 = \frac{B}{AeK}p.$$

According to this, when X varies as $\sqrt{p}$, B/AeK must be independent of the pressure p.

At low pressures with large currents the positive column becomes striated. No satisfactory theory of the striations has been worked out, but in some cases it appears that the potential difference between successive striations is about equal to the ionization potential of the gas, that is, the potential through which an electron must fall to acquire enough energy to ionize a gas molecule by colliding with it.

In a transverse magnetic field the positive column is deflected sideways like a flexible conductor carrying a current. If the positive column is in a tube of circular cross-section which it fills completely, a uniform transverse magnetic field causes a concentration of the luminosity on one side of the tube, so that the luminosity is greatest on one side and fades away gradually towards the other side. A transverse electric field or Hall Effect is produced by the magnetic field in the plane perpendicular to the magnetic field. This Hall Effect can be examined with the apparatus of fig. 2 (p. 336). The glass stopper is turned until the two small electrodes C and D are at the same potential so that they are in a plane perpendicular to the current. When a magnetic field is then produced parallel to the electrodes C and D a potential difference is produced between them which can be measured with an insulated quadrant electrometer. If Z denotes the transverse electric field, H the magnetic field, and p the gas pressure, then in air between $p = 0{\cdot}26$ mm. and $2{\cdot}9$ mm. the writer found $Z = 0{\cdot}0248\ H/p$. In hydrogen, $Z = 0{\cdot}0205\ H/p$, and in oxygen, $Z = 0{\cdot}0379\ H/p$.

The current is carried almost entirely by the negative electrons, so that the force on the electrons due to the magnetic field must be equal and opposite to that due to the transverse electric field Z. Hence, if v is the velocity of the electrons along the tube, we have

$$Hev = Ze,$$

or

$$v = \frac{Z}{H}.$$

Since $Z = A\frac{H}{p}$, where A is a constant, we get $v = A/p$, so that the velocity of the electrons along the positive column is inversely as the gas pressure. For air, $A = 0.025$ with Z in volts per centimetre, so that with Z in electromagnetic units $A = 25 \times 10^5$, and $v = \frac{25 \times 10^5}{p}$ cm. per second. At 1 mm. pressure in air $X = 35$ volts per centimetre, and $X/p = 35$. For $X/p = 35$, Townsend found for the velocity of electrons in air about 120×10^5 cm. per second, which is about five times that just deduced from the Hall Effect. This probably indicates that the theory of the Hall Effect just given needs modification. It is found that the Hall Effect in discharge tubes varies along the discharge in a similar way to the electric intensity. It is large in the Crookes dark space, and small in the negative glow and Faraday dark space.

The electrical conductivity of the discharge can be measured by measuring the current due to a small potential difference between two small electrodes a few millimetres apart. In this way it is found that the conductivity is small in the Crookes dark space and very large relatively in the negative glow. It is greater in the positive column than in the Faraday dark space. Roughly speaking, the conductivity is inversely as the electric intensity along the discharge, as we should expect.

REFERENCES

1. *Conduction of Electricity through Gases.* J. J. Thomson and G. P. Thomson.
2. *Electricity in Gases.* J. S. Townsend.

CHAPTER XVII

Atmospheric Electricity

1. Vertical Field in the Atmosphere. Conductivity.

Besides thunderstorms other less noticeable electrical phenomena are observed in the open air. There is usually a vertical electric field of the order of 150 volts per metre. The difference of potential between the earth and a point in the air above it may be found by means of an insulated conductor provided with some device to bring it to the same potential as the surrounding air. A small flame on the conductor, or some radium, may be used. The air close to the conductor is made conducting so that any charge on it leaks away and it is then at the same potential as the air near it. The potential difference between the conductor and the ground can be measured with an electrostatic voltmeter connected to the conductor and to the ground by insulated wires. If the conductor is on a pole 10 m. above the ground in the open air away from buildings or trees, the potential difference between it and the ground will be of the order of 1500 volts. The vertical field varies greatly. In fine dry weather it is usually directed downwards, indicating a negative charge on the earth's surface. It varies with the time of day and season of the year.

The vertical field has been measured at various heights by means of balloons. It is found to diminish as the height increases, and usually becomes negligible at about 10,000 m.

The air in the open is not a perfect insulator. It always contains positive and negative ions and so conducts to some extent. The conductivity can be measured by passing a stream of air through a metal case containing a charged insulated electrode, and measuring the rate at which the electrode loses its charge. Another very good method due to C. T. R. Wilson is to use a rather large horizontal metal plate placed level with the ground. The plate is insulated and connected to a capillary electrometer, which records photographically the total amount of electricity entering or leaving the plate. If the plate is covered with an earthed metal cover the charge on it is zero. If the cover is then removed, the vertical electric field induces a charge on the plate which is indicated by the electrometer, so that the strength of the field can be calculated. If the plate is left uncovered the electro-

meter indicates a small current usually flowing from the plate through the electrometer to the ground. If A is the area of the plate, F the strength of the vertical field, and C the current through the air to the plate, then $C = \mu AF$, where μ is the conductivity of the air. The conductivity is of the order of 10^{-4} in electrostatic units. This means that a field of 1·5 volts per centimetre gives a current of $\frac{1}{2} \times 10^{-6}$ electrostatic units per second or $\frac{1}{6} \times 10^{-15}$ amperes per square centimetre. If we suppose that there is a downward vertical field of 1·5 volts per centimetre all over the surface of the earth then the negative charge on the earth is about $2{\cdot}5 \times 10^{15}$ electrostatic units or one million coulombs. The total current into the earth from the air, if the conductivity were 10^{-4} electrostatic units everywhere and the vertical field 1·5 volts per centimetre, would be $2{\cdot}5 \times 10^{12}$ electrostatic units or 1000 amperes, since the area of the earth's surface is 5×10^{18} sq cm. The current into the earth would therefore be enough to discharge it completely in 1000 sec or about 17 min.

2. How is the Earth's Charge Maintained?

One of the problems of atmospheric electricity is to explain how it is that the earth remains charged. However the conductivity and the vertical field are not known over a sufficiently large fraction of the earth's surface for us to be sure that there is really a current of 1000 amperes going into the earth through the air. The current may be upwards over part of the surface and downwards over the rest.

There is good reason to believe that the upper regions of the atmosphere, where the pressure is low, are comparatively good conductors. It is found that electric waves are reflected from the upper regions as though there were a conducting layer at a height of about 50 Km. The earth and this conducting layer therefore form an enormous condenser, the capacity of which is about 10^{11} cm. or 100,000 microfarads. If this condenser were charged so that the charge on the earth was one million coulombs the potential difference would be 9×10^{6} volts, and the vertical field would be about 2 volts per centimetre. However, the vertical field becomes small at 10 Km., so that the actual potential difference is probably usually not more than one million volts.

Various suggestions have been made to explain how the charge on the earth is maintained. C. T. R. Wilson considers that rain drops carry down enough charge to the earth to keep up its charge. Rain is usually charged more or less, and according to C. T. R. Wilson more often with negative than with positive electricity. Some observers, however, say it is generally positively charged. When a lightning flash strikes the earth it must evidently give or take away a considerable amount of electricity. We might suppose that enough flashes strike the earth on the average to keep up its charge. Very interesting observations have been made by C. T. R. Wilson, using the insulated

plate apparatus described above. When a thunderstorm occurs anywhere near this apparatus each flash is recorded by the electrometer showing a sudden rush of electricity. This indicates a sudden change in the vertical field, which then gradually changes back to its normal value The change of field due to a discharge from a cloud to the earth can be easily calculated. Let a charge E be on a cloud at a height h above the earth. There will be an equal and opposite induced charge on the earth, and the field due to these charges is the same as that due to the charge E above the earth and a charge $-E$ at a depth h below the surface. The two charges therefore act like a doublet of moment $2Eh$, and the vertical field due to this doublet on the earth's surface at a distance r is equal to $2Eh/r^3$, provided r is large compared with h. When the cloud is discharged by a flash to the earth the vertical field therefore suddenly changes by $2Eh/r^3$, so that Eh can be calculated when r is known. C. T. R. Wilson got r by measuring the time between the flash and the thunder. It was found that the results obtained could be best explained by supposing the charge E to be about 30 coulombs, and the height h from 8 to 15 Km. More flashes in which the current was upwards than downwards were observed. To keep up the negative charge on the earth, assuming it is the same all over the earth, would therefore require about 30 more upward flashes per second than downward flashes on the average During thunderstorms the electric field is frequently directed upwards, indicating a negative charge on the clouds. A charge of 30 coulombs on a cloud is a reasonable amount of electricity. For example, a spherical cloud of radius 1 Km. with charge of 30 coulombs would have a field at its surface of only 3000 volts per centimetre, and an average density of charge of only about $2 \times 10^{-}$[illegible] electrostatic units per cubic centimetre. It seems to be considered that lightning flashes cannot be supposed to give more than a small fraction of the current of about 1000 amperes believed necessary to keep up the negative charge on the earth, but anyone who has seen a violent tropical thunderstorm might be inclined to doubt this opinion. For example, suppose we assume that over one-tenth of the surface of the earth there are on the average ten storms per year per hundred square miles. This gives two million storms per year A current of 1000 amperes gives about 3×10^{10} coulombs in a year, so that each storm would have to give about 2×10^{4} coulombs to the earth This would mean about 1000 flashes per storm, which seems a very low estimate. In tropical storms the flashes often follow each other so quickly that it is hard to count them, and the storm may remain in the vicinity of one place for several hours. It seems to the writer quite reasonable to suppose that lightning flashes on the average supply electricity to the earth at the rate of 1000 coulombs per second.

In thunderstorms there is usually a rapid upward current of air near the centre of the storm. The expansion of the air as it rises

causes supersaturation of the water vapour, which therefore condenses on any nuclei present. Negative ions act as nuclei more readily than positive ions, so that the drops formed will tend to be negatively charged. The drops formed fall slowly in the air, and the positive ions will therefore be carried up above the cloud into the upper conducting regions, leaving the cloud negatively charged. The positive charge which gets up to the conducting regions will be spread out to a considerable extent over the conducting layer and will induce a negative charge on the earth. If the negatively charged cloud falls far enough it may discharge to the earth either by a lightning flash or by merely carrying its charge down to the earth on the rain drops. In this way the earth gets a negative charge and the upper regions a positive charge After the cloud is discharged to the earth, the positive charge left in the upper regions will spread out over the whole area of the conducting layer and induce an equal negative charge over the whole surface of the earth. The positive charge of course need not be carried up 50 Km. by the storm; so long as it gets up well above the clouds it will induce a negative charge in the conducting regions above it and eventually be attracted to higher levels.

Various other suggestions as to how the earth is kept charged have been made. A theory due to Ebert relies on the fact that ionized air in contact with liquid or solid bodies acquires a positive charge owing to the negative ions diffusing more rapidly than the positive. Air at the earth's surface therefore gets positively charged, and Ebert suggested that this positively charged air is carried up by air currents. This theory is not unlike the thunderstorm theory. We may also suppose that rain drops get negatively charged by the diffusion of negative ions into them and that the positive ions are carried up by air currents. Positively charged rain is frequently observed, which is a fact against the thunderstorm and Ebert's theories.

Another theory due to G. C. Simpson is that positively and negatively charged particles are shot out from the sun and that some of these reach the earth. The positive particles are supposed to be positively charged atoms which are stopped in the upper regions of the atmosphere, while the negative particles are electrons which are supposed to penetrate the atmosphere and get into the earth. This requires the electrons to have velocities almost equal to the velocity of light, and it is difficult to see how such high-velocity electrons are produced. We might suppose that they start from the sun with small velocities corresponding to the sun's temperature and are accelerated by the pressure of the sun's radiation on them, so getting nearly up to the velocity of light before they reach the earth. Such electrons might be expected to ionize the air, but W. F. G. Swann has pointed out that electrons moving with almost the velocity of light produce very few ions. They do not remain long enough near an atom to do enough

work on the electrons in the atom to knock them out. Swann has tried to detect the charges carried by these electrons by means of a large insulated mass of copper. Some of the fast electrons ought to be absorbed by such a mass so that it ought to acquire a negative charge. No such charge, however, could be detected.

Another suggestion due to W. F. G. Swann is that the very penetrating cosmic rays investigated by Millikan which come into the atmosphere from outside produce penetrating β-rays which travel on mainly in the direction of the cosmic rays. These β rays would thus carry a negative charge along a certain distance towards the earth, and the production of such β-rays throughout the atmosphere all directed towards the earth would result in a flow of negative electricity into the earth from the atmosphere. If n such β-rays are produced per cubic centimetre per second in the atmosphere by the cosmic rays, and they all travel a distance l towards the earth, the current density due to them would be nle. The current density required to keep the earth charged is about $\frac{1}{2} \times 10^{-6}$ electrostatic units, so that since $e = 5 \times 10^{-10}$ we have $\frac{1}{2} \times 10^{-6} = nl \times 5 \times 10^{-10}$ or $nl = 10^3$. If then $n = 1$ we must have $l = 10^3$ cm. These β-rays would ionize the air, and the number of ions per cubic centimetre would be equal to n times the number of ions produced by one β-ray. This, however, is a large number, since β rays ionize strongly before they are stopped. It seems that this theory would give too much ionization and so will not do.

It will be seen from this discussion that we are as yet very far from having a satisfactory theory of atmospheric electricity. The facts available are not sufficient to enable definite conclusions to be reached. It is quite likely that when observations have been made all over the earth for a long time, it will appear that the observations now available were quite inadequate as a basis of a theory of the phenomena. The ionization of the atmosphere is attributed to radioactive radiations and to Millikan's cosmic rays. It is found that the air near the earth contains minute traces of radium emanation. The α- and β-rays from this emanation and its products are sufficient to account for most of the observed ionization The emanation is presumably evolved by traces of radium in the surface of the earth.

Reference

"The Mechanism of a Thunderstorm." G. C. Simpson, *Proc. Royal Soc.*, April, 1927.

CHAPTER XVIII

Special Relativity

1. Relativity in Newtonian Dynamics.

In ordinary physical experiments the building in which the experiments are made is regarded as at rest, and the positions and velocities of the bodies used are measured relatively to the building. The building is on the earth, so that we may say that the material frame of reference which is used and regarded as at rest is the earth.

All the material bodies in the building when not supported are found to have the same acceleration g vertically downwards. If, therefore, we take axes x, y, z with y vertically upwards and x and z in a horizontal plane, the equations of motion of a particle are

$$\begin{aligned} m\ddot{x} &= X, \\ m\ddot{y} &= -mg + Y, \\ m\ddot{z} &= Z, \end{aligned}$$

where x, y, z are the co-ordinates of the particle, and X, Y, Z are the components of any external force which may be applied to it; the axes being supposed to be fixed on the building. If $X = Z = 0$, and $Y = mg$, then $\ddot{x} = \ddot{y} = \ddot{z} = 0$, and the particle moves with constant velocity in a straight line.

If instead of using axes fixed on the building we use axes moving relatively to the building with constant velocity in a straight line, the equations of motion are unchanged.

For example, suppose we use axes x', y', z' fixed on an elevator in the building which is moving upwards with constant velocity v, and suppose that at time $t = 0$ the two sets of axes coincide exactly. Then at any later time $x = x'$, $z = z'$, but $y = y' + vt$. Substituting these values in the equations of motion, we get

$$\begin{aligned} m\ddot{x}' &= X, \\ m\ddot{y}' &= -mg + Y, \\ m\ddot{z}' &= Z, \end{aligned}$$

as before. If, however, we use axes moving relatively to the building with a variable velocity, the equations of the particle do not remain unchanged in form. For example, suppose we again use the axes

fixed on the elevator but that it moves upward with a constant acceleration a, then $x = x'$, $z = z'$ and $y = y' + \frac{1}{2}at^2$. Substituting these values in the equations of motion we get

$$m\ddot{x}' = X,$$
$$m\ddot{y}' = -m(g + a) + Y,$$
$$m\ddot{z}' = Z.$$

In this case if $X = 0$, $Z = 0$ and $Y = m(g + a)$, then $\ddot{x}' = \ddot{y}' = \ddot{z}' = 0$, and the particle moves with uniform velocity in a straight line. The additional force ma gives the particle an acceleration equal to that of the axes, and neutralizes the effect of the acceleration of the axes.

Let us suppose now that an observer on a material body anywhere, not necessarily on the earth, uses rectangular axes fixed on the body and observes the motion of a particle relatively to these axes. Suppose he finds the particle moves with constant velocity in a straight line when he applies no forces X, Y, Z to it and that in general its equations of motion are

$$m\ddot{x} = X, \quad m\ddot{y} = Y, \quad m\ddot{z} = Z.$$

What conclusions can he draw from this as to the motion of his axes through space? Is he entitled to conclude that his axes are not moving with an acceleration? As we have just seen, a motion of the axes with constant velocity in a straight line makes no difference, so it is clear that he cannot draw any conclusions as to the velocity of his system. Also, we have seen that when an acceleration is given to the axes its effect can be neutralized by applying a suitable force to the particle. It follows that the observer's axes may have any acceleration provided his system is in a field of force which neutralizes this acceleration. Thus, if his axes have an acceleration a in any direction, and if there is also present a field of force which gives to the particle an equal acceleration, then the equations of motion of the particle relative to his axes will be

$$m\ddot{x} = X, \quad m\ddot{y} = Y, \quad m\ddot{z} = Z.$$

If, for example, the observer's system is in a uniform gravitational field, then this field will give to his system and axes and to any particle he may use the same acceleration. Suppose this field is in the direction of his y axis, and that the acceleration it produces is f. Then if x', y', z' are the co-ordinates of a point in this system, and x, y, z the co-ordinates of the same point relative to axes at rest which coincide with x', y', z' at $t = 0$, we have $x = x'$, $y = y' + \frac{1}{2}ft^2$, $z = z'$, and the equations of motion relative to the fixed axes

$$m\ddot{x} = X, \quad m\ddot{y} = mf + Y, \quad m\ddot{z} = Z,$$

become

$$m\ddot{x}' = X, \quad m\ddot{y}' = Y, \quad m\ddot{z}' = Z,$$

relative to the accelerated axes.

It appears, therefore, that the observer cannot draw any conclusions as to the motion of his system through space from his observations on the motion of a particle in his system.

If the particle does not move with uniform velocity in a straight line when no forces are applied to it, this may be due either to a field of force or to a non-uniform motion of the axes, but the observer cannot tell which explanation is the correct one.

On the surface of the earth it is found that a particle, when no forces are applied to it, does not move with constant velocity but has an acceleration g vertically downwards. An observer in a laboratory may observe this acceleration g, but without other observations outside the laboratory he cannot tell whether it is due to an acceleration of the laboratory or to a field of force acting inside the laboratory or to a combination of the two. As a matter of fact, the acceleration g is believed to be partly due to the gravitational field of the earth and partly to the acceleration of the laboratory as it moves with the earth.

If there were a field of force in the laboratory which did not give the same acceleration to material particles made of different kinds of matter, such a field could be distinguished from the effect of an acceleration of the laboratory, since of course an acceleration of the laboratory must give the same opposite acceleration to any free particle in it. For example, an electric field gives no acceleration to an uncharged particle but gives one to a charged particle.

The only kind of field of force which gives the same acceleration to all particles of whatever kind is a gravitational field. We conclude, therefore, that observations on the motion of material particles can give no information as to the velocity or acceleration of the material frame of reference used, because the velocity makes no difference, and the effects of the acceleration cannot be distinguished from those of a gravitational field.

The fact that the equations of motion of a particle are unchanged by a uniform motion of the material frame of reference may be referred to as the special principle of mechanical or Newtonian relativity. According to this principle, if x, y, z, t are the co-ordinates of a particle in a system S, and x', y', z', t' the co-ordinates in another system S' moving along x with velocity v and coinciding with S at $t = t' = 0$, then $x = x' + vt$, $y = y'$, $z = z'$, $t = t'$ are the equations for transforming the equations of motion of any particle from S to S'.

The fact that when S' has any acceleration relative to S the equations of motion in S' are still the same as in S, provided a suitable gravitational field acts in S', may be referred to as the general principle of mechanical relativity.

2. **The Ether, or Space.**

In the study of solid geometry, the relations of the distances between different points on a solid are discussed without reference to the physical properties of the material of the solid, and in this way by a process of abstraction the idea of the geometrical properties of empty space has arisen. It has been supposed that empty space has no physical properties but only geometrical properties. No such empty space without physical properties has ever been observed, and the assumption that it can exist is without justification. It is convenient to ignore the physical properties of space when discussing its geometrical properties, but this ought not to have resulted in the belief in the possibility of the existence of empty space having only geometrical properties.

It is found that a vacuum, that is, a space from which all material particles have been removed, has important physical properties. Light waves can pass through it, and electric, magnetic, and gravitational actions can take place across it. It has specific inductive capacity and magnetic permeability. The existence of these physical properties in empty space, together with the conception of empty space having only geometrical properties, lead to the idea that space is filled with a medium, which has been called the ether, to which the physical properties in question belong. That is, we conceive really empty space as having only geometrical properties, and therefore, since we find that actual space has also physical properties, we suppose that actual space consists of our imaginary empty space filled with a medium having the observed physical properties. It is clear that there is no logical justification for this way of regarding the matter. We find that space containing no material particles has physical properties, and since we cannot separate this space in any conceivable way into two parts, one having the geometrical properties and the other the physical, we must regard the geometrical and physical properties as equally the properties of space.

If the physical properties of space could be removed, the geometrical properties would probably also disappear, in fact we are not really justified in drawing any distinction between these two sorts of properties; the geometrical properties of space are just as much physical properties as the specific inductive capacity and magnetic permeability. Descartes said that if everything inside a hollow vessel were removed its sides would be in contact. There would not even be space left in it. This profound remark shows that Descartes was not confused by an imaginary separation of the geometrical and physical properties of space.

The idea of an ether is therefore seen to be superfluous, since it is based on a purely imaginary separation of the geometrical and physical

properties of space. If it is desired to retain the word ether it may be done by using it to mean the actual space of experience as distinguished from the various imaginary types of space discussed in geometry In what follows we shall use the word space to mean the actual space of experience with its inseparable geometrical and physical properties

Material bodies are observed to move relatively to each other in space, and they are said to excite in the space around them fields of force which move with them. Also we have fields of force moving alone through space, without material particles moving with them. The distinction between material particles and their fields of force is not very clearly defined The physical properties of the particles are the properties of their fields, and it is not clear that a particle is anything more than the centre from which its field radiates. The momentum of matter for example is believed to be the electromagnetic momentum of its electric and magnetic fields.

The only kinds of motion of which we have any evidence are the motions of material particles and fields of force through space. When it was supposed that space was filled with ether which was imagined to be different from space, the possibility of this ether moving through space was discussed, but we need not now consider this as a possible form of motion since we do not now admit any distinction between ether and space.

3. Motion through Space—the Michelson-Morley Experiment.

We may, however, consider whether the motion of a material system through space can produce any effects observable on the system. In particular, is it possible to determine the motion of the earth through space by means of observations made in a laboratory on the earth? We have seen that observations on the motion of a material particle in a laboratory give no information as to the motion of the laboratory through space. If space had only geometrical properties, we should not expect motion through it to produce any observable effects, but since this is not the case there is no *a priori* reason why motion through it should not produce effects which could be detected. We have seen that no such effects on the motion of material particles are to be expected, but the possibility of electromagnetic and optical effects remains. Up to about the year 1925 all attempts to detect any such effects failed, and it came to be generally believed that this was due, not to the experiments tried being insufficiently sensitive, but to the nature of the relations between space and matter, which were supposed to be such that the determination of the motion of a system through space by means of observations on the system was impossible. This idea was adopted by Einstein as the basis of his theory of relativity, a theory of great interest which leads to highly important results. Since 1925, however, experiments by D. C. Miller on Mount Wilson in

California, by a method originally due to Michelson, of Chicago, seemed at first sight to show that possibly optical effects due to the motion of the earth through space could be detected, and that the magnitude and direction of the velocity of the earth through space could be measured by observations made in a laboratory on the earth. These experiments threw some doubt upon the general validity of the theory of relativity, which would have had to be abandoned if Miller's results had been confirmed. The effects apparently observed by Miller were extremely small, and were most likely merely due to errors of some unexpected kind. Recent very exact repetitions of Michelson's experiment have failed to confirm Miller's results.

Let us suppose that an observer on a material system, such as the earth, which is moving through space with a velocity v, makes observations on the velocity of light. Suppose he determines the time taken by light to pass from a point A to a point B at a distance d from A and to be reflected back to A. Also, let the velocity v of his system be in the direction from A to B. Let the light start from A at time t_1, arrive at B at t_2, and get back to A at t_3. Then if the light travels through space with velocity c, we have

$$d + v(t_2 - t_1) = c(t_2 - t_1),$$

$$d - v(t_3 - t_2) = c(t_3 - t_2),$$

or

$$t_2 - t_1 = \frac{d}{c - v},$$

$$t_3 - t_2 = \frac{d}{c + v},$$

so that

$$t_3 - t_1 = \frac{2d}{c(1 - v^2/c^2)}.$$

Suppose now that AB is at right angles to the velocity v, and that the mirror at B is arranged to reflect the light from A back to A as before. In this case A travels a distance $v(t_3 - t_1)$ at right angles to AB, so that the distance the light goes is $2\sqrt{d^2 + \frac{v^2}{4}(t_3 - t_1)^2}$, and we have

$$2\sqrt{d^2 + \frac{v^2}{4}(t_3 - t_1)^2} = c(t_3 - t_1),$$

or

$$t_3 - t_1 = \frac{2d}{c\sqrt{1 - v^2/c^2}}.$$

The interval $t_3 - t_1$ is therefore not the same when v is along AB as when v is at right angles to AB, so that the observer, by measuring the time for light to go a distance d from A and back in different direc-

tions in his system, could apparently determine the magnitude and direction of the velocity of his system through space.

We have assumed here that the velocity of light c through space is a constant independent of the velocity of the material system. The source of light used moves with the system, so that we have assumed that the velocity of the light emitted by a moving source is independent of the velocity of the source. These assumptions are in accordance with the electromagnetic wave theory of light. We suppose that the velocity of light through space is equal to $1/\sqrt{\mu K}$, where μ is the magnetic permeability and K the specific inductive capacity of space.

An experiment to compare the velocities of light in different directions was devised by Michelson, and carried out by him and Morley, and later by D. C. Miller.

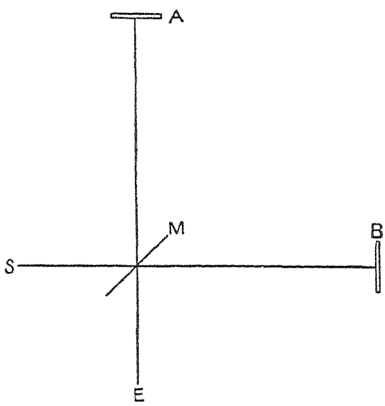


Fig 1

Light from a source S (fig. 1) falls at 45° on a glass plate M, where it is partly reflected to a mirror A and partly transmitted to another mirror B. The mirrors A and B reflect the light back to M, where it is again partly reflected and partly transmitted. Thus two beams arrive at E, one of which has gone along the path SMAME, and the other along the path SMBME. If the two paths to E are equal, and white light is used, a system of interference bands is seen at E. If one of the mirrors is moved towards or away from M, the bands move. Changing MA by one-half wave-length causes a dark band to move through the distance between two adjacent dark bands. Thus, by observing the interference bands seen at E, any change in the difference between the times taken by the light to traverse the two paths can be detected. The apparatus was mounted in a horizontal plane on a block of stone floating on mercury, so that it could be slowly rotated about a vertical axis and the interference bands could be observed during the rotation. The paths MB and MA would then rotate relatively to the horizontal component of the velocity of the earth through space, and when one was parallel to this velocity component the other would be perpendicular to it. We have seen that the time for light to go a distance d and back along the direction of the motion through space with velocity v is

$$\frac{2d}{c(1-v^2/c^2)}$$

while perpendicular to v it is

$$\frac{2d}{c\sqrt{1-v^2/c^2}},$$

so that we should expect a shift of the interference bands corresponding to a time difference

$$\frac{2d}{c}\left\{\frac{1}{1-v^2/c^2}-\frac{1}{\sqrt{1-v^2/c^2}}\right\}$$

during the rotation of the apparatus. The bands would oscillate about a mean position twice during each revolution of the apparatus. Michelson and Morley tried this experiment with great care, and concluded that there was no appreciable effect due to the orbital motion of the earth round the sun, although they estimated that they could have detected the effect due to a velocity of one-tenth of this orbital velocity. The time difference given above is approximately equal to v^2d/c^3 when v/c is small, and so corresponds to a path difference v^2d/c^2 which gives a shift of the bands equal to $v^2d/c^2\lambda$ times the distance from one band to the next one, where λ is the wave-length of the light used. The orbital velocity of the earth in its orbit round the sun is about 3×10^6 cm. per second, so that if $\lambda = 10^{-4}$ cm., we get

$$v^2d/c^2\lambda = 9 \times 10^{12}\,d/9 \times 10^{20} \times 10^{-4} = 10^{-4}d.$$

With $d = 10{,}000$ cm., which was about the value used by Michelson and Morley, $10^{-4}d = 1$, so that the shift expected was about equal to the distance between two bands. No such shift was observed, and this result became the principal foundation of the theory of relativity.

4. The Fitzgerald Contraction.

It was suggested by Fitzgerald and H. A. Lorentz that this negative result could be explained by supposing that material bodies contract slightly along the direction of motion when moving through space. Thus, in the Michelson-Morley experiment, if the distance along the velocity component v is d, and that perpendicular to the velocity d', then the time difference becomes

$$\frac{2}{c}\left\{\frac{d'}{\sqrt{1-v^2/c^2}}-\frac{d}{1-v^2/c^2}\right\},$$

which is zero if $$d = d'\sqrt{1-v^2/c^2}.$$

Such a contraction could not be detected by ordinary measurements, provided all bodies contract equally. The Michelson-Morley experi-

ment was repeated, different materials being used to fix the distances between the mirrors, with the same negative result.

We do not know the velocity of the earth through space. We can determine the velocity of the earth relative to the sun and stars, but these bodies may be moving through space with any velocity. Suppose for example that the earth were moving through space with a velocity $v = \frac{\sqrt{3}}{2}c$, then $\sqrt{1 - v^2/c^2} = \frac{1}{2}$, so that the length of a rod with its length perpendicular to the direction of the velocity would be double its length when parallel to the velocity.

Length therefore depends on the unknown velocity through space, and so has no absolute value, but is a relative quantity.

5. Einstein's Special Theory. Lorentz Transformation.

The negative result of the Michelson-Morley experiment is precisely what would have been obtained if the earth were at rest in space, and since no other effects due to the motion of the earth through space have been detected Einstein was led to put forward his *special* theory of relativity, according to which the motion of a material system through space with uniform velocity makes no difference to the phenomena observable in the system by observers on the system. This amounts to supposing that the principle of relativity, which as we have seen applies to the motion of material particles, can be extended to include electrical and optical phenomena as well.

Einstein some years later developed his *general* theory of relativity, according to which no motion of the material system could produce effects distinguishable from those due to gravitational fields either on optical phenomena or on the motion of material particles.

According to the special theory of relativity, the velocity of light should be the same in all directions to an observer on a system moving with any uniform velocity through space. Consider two such material systems S and S', and let the velocity of S' relative to S be v. Let position and time in S be measured by co-ordinates x, y, z, t and in S' by x', y', z', t'.

Let the two sets of axes coincide at $t = t' = 0$, and let the relative velocity v be along the axes x and x', so that to an observer on S the position of the origin of the co-ordinates x', y', z', t' in S' is given by $x = vt$, and to an observer in S' the position of the origin of x, y, z, t is given by the equation $x' = -vt'$.

Suppose that at the time $t = t' = 0$, when the two sets of axes coincide, a light wave is started at the origin of co-ordinates. To an observer on S this wave will be a sphere given by the equation $x^2 + y^2 + z^2 = c^2t^2$, and to an observer on S' it will be a sphere given by $x'^2 + y'^2 + z'^2 = c^2t'^2$.

There must therefore be such relations between x, y, z, t and x', y', z', t' that $x^2 + y^2 + z^2 - c^2t^2$ is equal to $x'^2 + y'^2 + z'^2 - c^2t'^2$. To the observer on S a point at rest in S having co-ordinates x, y, z relative to the S axes will have co-ordinates $x - vt$, y, z relative to the S' axes since the origin of the S' axes is at $x = vt$, $y = 0$, $z = 0$.

To this observer on S a measuring rod, used by an observer on S' and at rest in S', will be moving along the x direction with velocity v, so that, assuming the Fitzgerald-Lorentz contraction, when the rod is parallel to the x axis it will be contracted, and the units of length marked on it will be lengths $\sqrt{1 - v^2/c^2}$ instead of unity.

The observer on S will therefore consider the x' lengths as measured by the observer on S' to be measured in terms of a length $\sqrt{1 - v^2/c^2}$ as unit, so that the x' co-ordinate of the point in question will not be $x - vt$ when measured by the observer in S', but $(x - vt)/\sqrt{1 - v^2/c^2}$. Hence

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \quad y' = y, \quad z' = z.$$

If we substitute these values in $x'^2 + y'^2 + z'^2 - c^2t'^2$, we get

$$\frac{x^2 - 2xvt + v^2t^2}{1 - v^2/c^2} + y^2 + z^2 - c^2t'^2,$$

which is equal to $x^2 + y^2 + z^2 - c^2t^2$ if

$$t' = \frac{t - \dfrac{v}{c^2}x}{\sqrt{1 - v^2/c^2}}.$$

Thus, according to the special principle of relativity, we have

$$y' = y, \quad z' = z, \quad x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \quad \text{and} \quad t' = \frac{t - \dfrac{vx}{c^2}}{\sqrt{1 - v^2/c^2}}.$$

In the same way, to an observer in S' a unit length along x at rest in S will be contracted to a length $\sqrt{1 - v^2/c^2}$, so that a point in S' at x', y', z' will have S co-ordinates given by

$$x = \frac{x' + vt'}{\sqrt{1 - v^2/c^2}}, \quad y = y', \quad z = z';$$

and this requires that

$$t = \frac{t' + \dfrac{vx'}{c^2}}{\sqrt{1 - v^2/c^2}}.$$

These equations giving x, y, z, t in terms of x', y', z', t' agree with those giving x', y', z', t' in terms of x, y, z, t.

6. Relativity of Time.

It appears that, on a material system, time as well as length is related to the velocity of the system through space, so that since the velocity through space is unknown absolute time cannot be determined In Newtonian or classical dynamics, time was considered to be independent of position, but just how the relation between the time of an event at one place and the time of another event at another place was to be determined was not stated.

Consider a material system S moving through space with an unknown but uniform velocity, and let x, y, z, t be the space-time co-ordinates used by an observer on this system. Suppose the observer has a clock at the origin and another precisely similar clock at some other point How can he determine the relation between the times indicated by the two clocks? Suppose a light signal sent out from the origin at a time t_1 by the clock at the origin, received at the other clock at t_2, and reflected back to the origin and received there at t_3. If the system were at rest in space we should then have $t_2 = \frac{t_1 + t_3}{2}$. But according to the special principle of relativity motion through space with uniform velocity makes no difference, so Einstein considers that the relation between the times is given by $t_2 = \frac{t_1 + t_3}{2}$ in any case, whatever the velocity of the system through space may be. That is to say, we define the relation between the times at the two clocks by the equation $t_2 = \frac{t_1 + t_3}{2}$. To an observer on the material system, light travels with the same velocity in all directions, so that t_2 is half-way between t_1 and t_3.

In another system S' moving with uniform velocity v along the x axis of the system S and coinciding with S at $t = t' = 0$, an observer on S' would determine the relation between clocks in different positions in the same way as the observer in S by means of the equation $t_2' = \frac{t_1' + t_3'}{2}$. The times so determined in S' would not agree with the times in S' as observed by the observer in S. To the observer in S the system S' would be moving with velocity v, so that the time $t_2' - t_1'$ for a light signal to go from the origin of S' to a point $(x', 0, 0)$ would not be x'/c, but would be given by $t_2' - t_1' = \frac{x'}{c - v}$, and the time $t_3' - t_2'$ for the light to go back to the origin of S' would be given by

$t_3' - t_2' = \frac{x'}{c+v}$. These equations give $t_2' = \frac{t_3' + t_1'}{2} + \frac{v}{2c}(t_3' - t_1')$. Thus the assumption of the observer in S' that $t_2' = \frac{t_3' + t_1'}{2}$ would appear to the observer on S to be in error by $\frac{v}{2c}(t_3' - t_1')$.

The observer on S' would consider an event happening at his origin at the time $\frac{t_1' + t_3'}{2}$ as simultaneous with an event at $(x', 0, 0)$ at the time t_2', but the observer on S would consider these events not simultaneous but separated by a time interval $\frac{v}{2c}(t_3' - t_1')$ measured in terms of the unit of time used on S'.

The assumption that $t_2 = \frac{t_1 + t_3}{2}$ in all cases is equivalent to the assumption that light has the same velocity in all directions to an observer in any system, and, therefore, together with the Fitzgerald-Lorentz contraction, leads to the equations for x', y', z', t' in terms of x, y, z, t obtained above by making $x^2 + y^2 + z^2 - c^2t^2$ equal to $x'^2 + y'^2 + z'^2 - c^2t'^2$.

7. Composition of Velocities.

The assumption that the velocity of light is the same in all directions on a system moving through space with any uniform velocity is equivalent to assuming that the vector sum of the velocity of light and any other velocity is equal to the velocity of light.

The usual rule for the composition of velocities is therefore not correct according to the special theory of relativity. Suppose we have a particle moving along the x axis with a velocity v in a system S, and another particle moving along the x axis with velocity u. We may define the difference between v and u as the velocity of the first particle to an observer moving with the second particle. Let the observer then be on a system S' moving with velocity u relatively to S. The second particle is then at rest in S', and we may suppose that it is at the origin of S'. For the first particle let $x = vt$, so that for this particle in S', since the velocity of S relative to S' is $-u$, we have

$$x = \frac{x' + ut'}{\sqrt{1 - u^2/c^2}}$$

and

$$t = \frac{t' + ux'/c^2}{\sqrt{1 - u^2/c^2}}.$$

Substituting these values in $x = vt$, we get

$$\frac{x'+ut'}{\sqrt{1-u^2/c^2}} = \frac{vt'+\dfrac{vux'}{c^2}}{\sqrt{1-u^2/c^2}},$$

or

$$\frac{x'}{t'} = \frac{v-u}{1-uv/c^2}.$$

But x'/t' is the velocity of the first particle to the observer in S', that is, by the definition adopted, the difference between v and u.

In the same way the sum of the two velocities u and v is equal to $\dfrac{v+u}{1+uv/c^2}$.

If $u=c$ this becomes $\dfrac{v+c}{1+v/c}=c$, in agreement with the original assumption that the resultant of any velocity v and the velocity of light c is equal to c.

8. **Invariance in Expression of Physical Laws. Fizeau's Experiment.**

The assumption that motion of the material system through space with any uniform velocity makes no observable difference to phenomena on the system may be expressed in another way by saying that the mathematical equations expressing physical laws in terms of the co-ordinates x, y, z, t in a system S must be identical with the equations expressing the same laws in terms of the co-ordinates x', y', z', t' in a system S' moving with any uniform velocity relative to S. The equations for S must transform into identical equations for S' when the values of x, y, z, t in terms of x', y', z', t' are substituted in them.

This principle enables the effect of the motion of matter with uniform velocity on phenomena to be determined. For example, consider the motion of light through a medium having refractive index μ. The velocity of the light is c/μ when the medium is at rest, so that, if the light is moving along the x axis in a system S in which the medium is at rest then we have $x = ct/\mu$, where x is the distance the light has travelled from the origin in the time t.

Now suppose the light is observed by an observer in another system S' moving relative to S with velocity $-v$ along the x axis. We have

$$x = \frac{x'-vt'}{\sqrt{1-v^2/c^2}},$$

$$t = \frac{t'-vx'/c^2}{\sqrt{1-v^2/c^2}},$$

so that the equation $x = ct/\mu$ becomes

$$x' - vt' = \frac{c}{\mu}(t' - vx'/c^2)$$

in the system S'. Hence

$$\frac{x'}{t'} = \frac{c/\mu + v}{1 + v/\mu c}.$$

The velocity of light in the medium moving with velocity v is therefore increased by the motion of the medium from c/μ to $\left(\frac{c}{\mu} + v\right)/(1 + v/\mu c)$. When v/c is very small this is approximately equal to

$$\frac{c}{\mu} + v\left(1 - \frac{1}{\mu^2}\right).$$

This result agrees with the experimental results of Fizeau and Michelson obtained many years before the principle of relativity was developed.

If in a system S a medium of refractive index μ is moving along the x axis with velocity v, the velocity of light along the x axis in this medium is $\left(\frac{c}{\mu} + v\right)\Big/\left(1 + \frac{v}{\mu c}\right)$. The velocity of the light relative to the medium is then its velocity to an observer moving with the medium, or the difference between the two velocities, which is

$$\frac{\dfrac{c/\mu + v}{1 + v/\mu c} - v}{1 - \left(\dfrac{c/\mu + v}{1 + v/\mu c}\right)\dfrac{v}{c^2}} = \frac{c}{\mu},$$

according to the relativity formula for the composition of velocities.

Minkowski's Theory

9. An interesting way of regarding the special principle of relativity is due to Minkowski. The phenomena which we can observe on any material system S are a succession of events, each event taking place at a definite point x, y, z and at a definite time t. An event therefore has four co-ordinates x, y, z, t. If the same event is observed on another material system S' moving relative to S with uniform velocity v along the x axis and coinciding with S at $t = t' = 0$, its co-ordinates in S' will be x', y', z', t' and these are related to x, y, z, t in such a way that $x^2 + y^2 + z^2 - c^2t^2$ is equal to $x'^2 + y'^2 + z'^2 - c^2t'^2$.

Let $\tau = ict$ and $\tau' = ict'$, where $i = \sqrt{-1}$, so that $x^2 + y^2 + z^2 - c^2t^2$ becomes $x^2 + y^2 + z^2 + \tau^2$, and $x'^2 + y'^2 + z'^2 - c^2t'^2$ becomes

$x'^2 + y'^2 + z'^2 + \tau'^2$. Then if we take x, y, z, τ to be the rectangular co-ordinates of a point in a space of four dimensions, $\sqrt{x^2 + y^2 + z^2 + \tau^2}$ is equal to the distance of the point from the origin.

If we use another set of rectangular co-ordinates x', y', z', τ' having the same origin but inclined to x, y, z, τ, that is, rotate the axes x, y, z, τ into a new position, the distance of the point (x, y, z, τ) from the origin in terms of the new co-ordinates will be $\sqrt{x'^2 + y'^2 + z'^2 + \tau'^2}$, and of course this distance is unchanged by the change of axes.

In the same way, the distance between any two points (x_1, y_1, z_1, τ_1) and (x_2, y_2, z_2, τ_2) in the four-dimensional space is equal to

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 + (\tau_2 - \tau_1)^2},$$

and this becomes

$$\sqrt{(x_2' - x_1')^2 + (y_2' - y_1')^2 + (z_2' - z_1')^2 + (\tau_2' - \tau_1')^2}$$

when the axes are changed, but its magnitude is unchanged. If we denote the distance between two points in the four-dimensional space by s, then

$$s^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 + (\tau_2 - \tau_1)^2,$$

and $$ds^2 = dx^2 + dy^2 + dz^2 + d\tau^2,$$

or $$ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2,$$

where ds is the distance between two points very near together, so that $x_2 - x_1 = dx$, &c. s may be called the *interval* between the two events, and it is the same for all sets of rectangular axes in the four-dimensional space.

If we change from co-ordinates x, y, z, τ to x', y', z', τ' by rotating the axes in the plane (x, τ) through an angle θ from τ towards x, then

$$\begin{aligned} x' &= x\cos\theta - \tau\sin\theta, \\ \tau' &= \tau\cos\theta + x\sin\theta, \\ y' &= y, \\ z' &= z. \end{aligned}$$

Comparing with the equations

$$\begin{aligned} x' &= \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \\ t' &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}, \\ y' &= y, \\ z' &= z. \end{aligned}$$

we see that $\cos\theta = \frac{1}{\sqrt{1 - v^2/c^2}}$, and $\tau \sin\theta = \frac{vt}{\sqrt{1 - v^2/c^2}}$, or, since $\tau = ict$,

$$\sin\theta = \frac{-iv}{c\sqrt{1 - v^2/c^2}}.$$

Thus we may say that changing from a system S with axes x, y, z, t to a system S' moving relative to S along x with velocity v, with axes x', y', z', t' coinciding with x, y, z, t at $t = t' = 0$, is equivalent to rotating the axes x, y, z, τ through an angle θ in the (x, τ) plane.

The cosine of θ is greater than unity, so that θ is an imaginary angle.

The four-dimensional space is sometimes called the Minkowski world

If a number of events are represented by points in this space, the configuration of these points will be fixed by the intervals s between every pair of them, and so will be the same whatever rectangular axes like x, y, z, τ are used. The axes can be taken in any direction in the space. For example, if we draw the time axis perpendicular to the plane defined by any three points or events, these three events will be simultaneous in the resulting co-ordinates. In this way we see clearly the relative character of space and time, but it appears that there are absolute quantities connected with events, namely the intervals s between any two of them.

A curve drawn in the four-dimensional space is called a world line, and the world line of a particle represents the relation between its position and the time in any set of co-ordinates which may be selected. The shape of the world line is the same for any set of co-ordinates.

Since the configuration of the points representing events in the four-dimensional space represents what happens in the system considered, and is independent of the co-ordinate system used, the laws according to which the events happen should be capable of being expressed in a form independent of the co-ordinate system chosen. Any vector quantity should be capable of representation by a line drawn in the four-dimensional space, and so should have four components in any particular system of co-ordinates x, y, z, τ. The length of the line representing the vector quantity will be the same for all co-ordinate systems, and in any co-ordinate system its direction will be related in the same way to the world lines and points which represent the events in the four-dimensional space. The interval s between two events is an example of such a four-dimensional vector.

The components of such a vector will be obtained by multiplying its length by the cosines of the angles it makes with the axes. Let α_1, α_2, α_3, α_4 be these cosines for an element ds of a world line. The

components of ds are dx, dy, dz, and $d\tau$, so that $\alpha_1 = dx/ds$, $\alpha_2 = dy/ds$, $\alpha_3 = dz/ds$, and $\alpha_4 = d\tau/ds$. But $ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2$, and

$$v^2 = \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2,$$

so that $\left(\frac{ds}{dt}\right)^2 = v^2 - c^2$, or $\frac{dt}{ds} = \frac{i}{\sqrt{c^2 - v^2}}$, where $i = \sqrt{-1}$.

Therefore
$$\alpha_1 = \frac{dx}{ds} = \frac{dx}{dt}\frac{dt}{ds} = \frac{iv_x}{\sqrt{c^2 - v^2}},$$

and in the same way

$$\alpha_2 = \frac{iv_y}{\sqrt{c^2 - v^2}}, \quad \alpha_3 = \frac{iv_z}{\sqrt{c^2 - v^2}},$$

and
$$\alpha_4 = \frac{d\tau}{ds} = ic\frac{dt}{ds} = \frac{c}{\sqrt{c^2 - v^2}}.$$

In this way we can express the direction cosines of ds in terms of the velocity v corresponding to any particular set of co-ordinates x, y, z, τ.

Any vector V transforms like ds, so that

$$\begin{aligned} V^2 &= V_x^2 + V_y^2 + V_z^2 + V_\tau^2 \\ &= V_{x'}^2 + V_{y'}^2 + V_{z'}^2 + V_{\tau'}^2, \end{aligned}$$

where $V_{x'}$, $V_{y'}$, $V_{z'}$, $V_{\tau'}$ are the components of V along the axes x', y', z', τ'. Let $V_\tau = iV_t$, and then

$$\begin{aligned} V^2 &= V_x^2 + V_y^2 + V_z^2 - V_t^2 \\ &= V_{x'}^2 + V_{y'}^2 + V_{z'}^2 - V_{t'}^2. \end{aligned}$$

V_t may be called the time component of V, while V_τ is the imaginary time component. For a rotation in the x, t plane, or a change from axes x, y, z, t to axes x', y', z', t' moving along Ox with velocity v and coinciding with x, y, z, t at $t = 0$,

$$V_{x'} = \frac{V_x - \frac{v}{c}V_t}{\sqrt{1 - v^2/c^2}}, \qquad V_{y'} = V_y$$

$$V_{t'} = \frac{V_t - \frac{v}{c}V_x}{\sqrt{1 - v^2/c^2}}, \qquad V_{z'} = V_z$$

just as

$$x' = \frac{x - \frac{v}{c}ct}{\sqrt{1 - v^2/c^2}}, \qquad y' = y$$

and
$$ct' = \frac{ct - \frac{v}{c}x}{\sqrt{1 - v^2/c^2}}, \qquad z' = z.$$

These equations give

$$V_{x'}^2 + V_{y'}^2 + V_{z'}^2 - V_{t'}^2 = \frac{\left(V_x - \frac{v}{c}V_t\right)^2}{1 - v^2/c^2} + V_y^2 + V_z^2 - \frac{\left(V_t - \frac{v}{c}V_x\right)^2}{1 - v^2/c^2}$$
$$= V_x^2 + V_y^2 + V_z^2 - V_t^2.$$

10. Minkowski Velocity.

Let us now consider velocity in the Minkowski world. We wish to find a four-dimensional vector which is independent of the particular co-ordinates chosen, and which may be taken to represent this velocity. Consider an element ds of the world line of a particle, and let the τ axis be drawn along ds. In this case $ds = d\tau = ic\,dt$, so that $ds/dt = ic$ is the velocity of the particle along its world line. Let us suppose then that the Minkowski velocity of a particle is equal to ic in any co-ordinates, and directed along ds. When the τ axis does not coincide with ds then the velocity ic has components $ic\alpha_1$, $ic\alpha_2$, $ic\alpha_3$, and $ic\alpha_4$, which are equal to

$$\frac{v_x}{\sqrt{1 - v^2/c^2}}, \quad \frac{v_y}{\sqrt{1 - v^2/c^2}}, \quad \frac{v_z}{\sqrt{1 - v^2/c^2}}, \text{ and } \frac{ic}{\sqrt{1 - v^2/c^2}}.$$

When v/c is very small, these are approximately

$$v_x, \; v_y, \; v_z, \text{ and } ic.$$

Thus the x, y, z components of the Minkowski velocity ic are equal to the corresponding components of the three-dimensional velocity v when v/c is very small.

The time component of the Minkowski velocity is given by $v_\tau = iv_t$, so that since $v_\tau = ic/\sqrt{1 - v^2/c^2}$ we have $v_t = c/\sqrt{1 - v^2/c^2}$. If we denote the components of the Minkowski velocity by $v_{mx}, v_{my}, v_{mz}, v_{mt}$, then

$$v_{mx}^2 + v_{my}^2 + v_{mz}^2 - v_{mt}^2 = -c^2,$$

since

$$\frac{v_x^2 + v_y^2 + v_z^2 - c^2}{1 - v^2/c^2} = \frac{v^2 - c^2}{c^2 - v^2}c^2 = -c^2.$$

For a rotation of the axes in the x, τ plane we have, as for any four-dimensional vector,

$$v_{mx}' = \frac{v_{mx} - \frac{v}{c}v_{mt}}{\sqrt{1 - v^2/c^2}},$$

$$v_{mt}' = \frac{v_{mt} - \frac{v}{c}v_{mx}}{\sqrt{1 - v^2/c^2}},$$

$$v_{my}' = v_{my}, \quad v_{mz}' = v_{mz},$$

where v is the velocity of the origin of the x', y', z' axes along Ox.

11. Minkowski Force.

Momentum is the product of mass and velocity, so that corresponding to the Minkowski velocity we have a four-dimensional vector Minkowski momentum. If m_0 is the mass of a particle at rest, then its velocity components are 0, 0, 0, ic, so we take its momentum to have components 0, 0, 0, icm_0. If $\vec{M}_0$ is its momentum when at rest, then $\vec{M}_0 = \vec{icm_0}$. If this equation is true for the axes x, y, z, t in which it is at rest, then it will be true in axes x', y', z', t' moving with any uniform velocity relative to x, y, z, t, so we have $\vec{M} = \vec{icm_0}$, where $\vec{M}$ is the Minkowski momentum of the particle moving with any velocity v, and $\vec{ic}$ is the Minkowski velocity of the particle. Hence $M_x = m_0 v_{mx} = \frac{m_0 v_x}{\sqrt{1 - v^2/c^2}}$, with similar equations for M_y and M_z, and $M_t = \frac{m_0 c}{\sqrt{1 - v^2/c^2}}$. If we define the mass m of the moving particle by the equation $M_x = mv_x$, as usual, then we have $m = m_0/\sqrt{1 - v^2/c^2}$. The same result was obtained in Chapter I from electromagnetic theory.

Force is defined classically as rate of change of momentum, so we may try to obtain a Minkowski force or four-dimensional vector force from the rate of change of the Minkowski momentum. If V is any four-dimensional vector with components V_x, V_y, V_z, and $V_\tau = iV_t$, then δV will also be a four-dimensional vector with components δV_x, δV_y, δV_z, and $\delta V_\tau = i\delta V_t$. However, $\delta V/\delta t$ is not a four-dimensional vector because δt changes when the axes are changed. We have $\delta s^2 = \delta x^2 + \delta y^2 + \delta z^2 - c^2\delta t^2$, and δs does not change. Hence $(\delta s/\delta t)^2 = v^2 - c^2$, so that $-\delta s^2 = (c^2 - v^2)\delta t^2$ does not change. It follows that $\delta t\sqrt{1 - v^2/c^2}$ does not change with change of axes, and so $\delta V/\delta t\sqrt{1 - v^2/c^2}$ must be a four-dimensional vector.

If, then, $\vec{\delta M}$ is a small change of the Minkowski momentum, $\vec{\delta M}/\sqrt{1 - v^2/c^2}\,\delta t$ will be a four-dimensional vector, and so we may take the Minkowski force $\vec{F}$, say, to be given by

$$\vec{F} = \frac{1}{\sqrt{1 - v^2/c^2}} \frac{d\vec{M}}{dt}.$$

This reduces to the classical expression $\vec{F} = \frac{d\vec{M}}{dt}$ when v/c is small. We have then for the components of the Minkowski force

$$F_x = \frac{1}{\sqrt{1 - v^2/c^2}} \frac{d}{dt}\left(\frac{m_0 v_x}{\sqrt{1 - v^2/c^2}}\right),$$

with similar expressions for F_y and F_z, and

$$F_t = \frac{1}{\sqrt{1-v^2/c^2}} \frac{d}{dt}\left(\frac{m_0 c}{\sqrt{1-v^2/c^2}}\right).$$

12. Work and Energy.

We have assumed so far that the rest mass m_0 is a constant.

For a mass remaining at rest $F_t = c\dfrac{dm_0}{dt}$. The time component of the velocity $v_{mt} = c$, so that $F_t v_{mt} = c^2 \dfrac{dm_0}{dt}$. This may be interpreted as meaning that the mass m_0 increases at a rate equal to the rate of working of the force F_t divided by c^2. If we suppose m_0 initially equal to zero, then $m_0 c^2$ is equal to the work done by the force F_t. According to this the mass m_0 has energy $m_0 c^2$ or an amount of energy E has mass equal to E/c^2.

This relation between mass and energy was first obtained by Einstein from the theory of relativity. It also follows from electromagnetic theory as was shown in Chapter I. It appears that conservation of energy and conservation of mass are identical principles. The result that $E = mc^2$ has been experimentally confirmed in the study of nuclear reactions. It is found that the energy released in a nuclear reaction is equal to the loss of total mass multiplied by the square of the velocity of light.

13. World Tensors.

Consider any two four-dimensional vectors p and q, having components p_x, p_y, p_z, $p_\tau = ip_t$, and q_x, q_y, q_z, $q_\tau = iq_t$. We can form sixteen products of one component of p with one of q thus:

$$\begin{array}{cccc} p_x q_x & p_x q_y & p_x q_z & ip_x q_t \\ p_y q_x & p_y q_y & p_y q_z & ip_y q_t \\ p_z q_x & p_z q_y & p_z q_z & ip_z q_t \\ ip_t q_x & ip_t q_y & ip_t q_z & -p_t q_t. \end{array}$$

These sixteen products are said to be the sixteen components of a world tensor. In general a world tensor is defined as a quantity having sixteen components which transform from one set of axes x, y, z, τ to any other set in the same way as the sixteen products of the components of a pair of vectors.

If t denotes a world tensor, its components may be denoted by t with two suffixes, as, for example, t_{xy} or $t_{y\tau}$.

If interchanging the two suffixes does not change the value of any component so that, for example, $t_{xy} = t_{yx}$ and $t_{y\tau} = t_{\tau y}$, the tensor is said to be symmetrical, and can be specified by ten quantities.

If interchanging the two suffixes of any component merely changes the sign of the component so that, for example, $t_{xy} = -t_{yx}$ and $t_{x\tau} = -t_{\tau x}$, then t_{xx}, t_{yy}, t_{zz} and $t_{\tau\tau}$ are all zero, and the tensor is said to be skew-symmetrical and can be specified by only six quantities, viz. t_{xy}, t_{xz}, t_{yz}, $t_{x\tau}$, $t_{y\tau}$, $t_{z\tau}$.

The vector divergence $\mathbf{V}$ of a world tensor is defined by the equations giving its four components. The x component is given by

$$V_x = \frac{\partial t_{xx}}{\partial x} + \frac{\partial t_{xy}}{\partial y} + \frac{\partial t_{xz}}{\partial z} + \frac{\partial t_{xt}}{c\partial t},$$

where $\dfrac{\partial t_{xt}}{c\partial t} = \dfrac{\partial t_{x\tau}}{\partial \tau}$, since $d\tau = icdt$ and $t_{x\tau} = it_{xt}$, with similar equations for V_y, V_z, and $V_\tau = iV_t$.

The components of the vector divergence of a skew-symmetrical tensor are then given by the equations

$$V_x = \frac{\partial t_{xy}}{\partial y} + \frac{\partial t_{xz}}{\partial z} + \frac{\partial t_{xt}}{c\partial t},$$

$$V_y = \frac{\partial t_{yx}}{\partial x} + \frac{\partial t_{yz}}{\partial z} + \frac{\partial t_{yt}}{c\partial t},$$

$$V_z = \frac{\partial t_{zx}}{\partial x} + \frac{\partial t_{zy}}{\partial y} + \frac{\partial t_{zt}}{c\partial t},$$

$$V_t = \frac{\partial t_{tx}}{\partial x} + \frac{\partial t_{ty}}{\partial y} + \frac{\partial t_{tz}}{\partial z}.$$

Now let $H_x = t_{yz}$, $H_y = t_{zx}$, and $H_z = t_{xy}$; and also let $E_x = t_{tx}$, $E_y = t_{ty}$, $E_z = t_{tz}$; so that

$$V_x = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - \frac{1}{c}\frac{\partial E_x}{\partial t},$$

$$V_y = \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} - \frac{1}{c}\frac{\partial E_y}{\partial t},$$

$$V_z = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - \frac{1}{c}\frac{\partial E_z}{\partial t},$$

$$V_t = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z},$$

since $t_{yz} = -t_{zy}$, $t_{tz} = -t_{zt}$, &c.

Thus if $\mathbf{V}_1$ denotes the component of $\mathbf{V}$ at right angles to the τ axis, so that $V^2 = V_1^2 + V_\tau^2$ and $V_1^2 = V_x^2 + V_y^2 + V_z^2$, and if $\mathbf{E}$, $\mathbf{H}$ denote the three-dimensional vectors of which the components are E_x, E_y, E_z, and H_x, H_y, H_z, we have

$$\mathbf{V}_1 + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \text{curl}\,\mathbf{H}$$

and
$$V_t = \text{div}\,\mathbf{E}.$$

In the four-dimensional space the curl or rotation of a vector may be defined just as in three-dimensional space. Thus the components of the curl R of a vector **P** having components P_x, P_y, P_z, $P_\tau = iP_t$ are defined by the equations

$$R_{xy} = \frac{\partial P_y}{\partial x} - \frac{\partial P_x}{\partial y}, \quad R_{yz} = \frac{\partial P_z}{\partial y} - \frac{\partial P_y}{\partial z},$$
$$R_{zx} = \frac{\partial P_x}{\partial z} - \frac{\partial P_z}{\partial x}, \quad R_{x\tau} = \frac{\partial P_\tau}{\partial x} - \frac{\partial P_x}{\partial \tau},$$
$$R_{y\tau} = \frac{\partial P_\tau}{\partial y} - \frac{\partial P_y}{\partial \tau}, \quad R_{z\tau} = \frac{\partial P_\tau}{\partial z} - \frac{\partial P_z}{\partial \tau}.$$

Differentiating R_{xy} with respect to z, R_{yz} with respect to x, and R_{zx} with respect to y, and adding the three equations so obtained, we get

$$\frac{\partial R_{yz}}{\partial x} + \frac{\partial R_{zx}}{\partial y} + \frac{\partial R_{xy}}{\partial z} = 0.$$

In the same way we get

$$\frac{\partial R_{z\tau}}{\partial y} + \frac{\partial R_{\tau y}}{\partial z} + \frac{\partial R_{yz}}{\partial \tau} = 0,$$
$$\frac{\partial R_{\tau x}}{\partial z} + \frac{\partial R_{xz}}{\partial \tau} + \frac{\partial R_{z\tau}}{\partial x} = 0,$$
$$\frac{\partial R_{xy}}{\partial \tau} + \frac{\partial R_{y\tau}}{\partial x} + \frac{\partial R_{\tau x}}{\partial y} = 0.$$

We have $R_{xy} = -R_{yx}$, &c., so that the six components of the curl R may be regarded as the components of a skew-symmetrical world tensor. Thus when a world tensor is the curl of a vector it must be a skew-symmetrical tensor, and its components must satisfy the four equations just obtained.

Let us now suppose that the world tensor R is identical with the tensor h previously considered, so that

$$R_{yz} = H_x, \quad R_{zx} = H_y, \quad R_{xy} = H_z,$$
$$R_{tx} = E_x, \quad R_{ty} = E_y, \quad R_{tz} = E_z,$$

where
$$R_{\tau x} = iR_{tx}, \; R_{\tau y} = iR_{ty}, \; R_{\tau z} = iR_{tz}.$$

Substituting these values we obtain

$$\frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} = 0,$$

$$-\frac{\partial E_z}{\partial y} + \frac{\partial E_y}{\partial z} - \frac{1}{c}\frac{\partial H_x}{\partial t} = 0,$$

$$\frac{\partial E_x}{\partial z} + \frac{1}{c}\frac{\partial H_y}{\partial t} - \frac{\partial E_z}{\partial x} = 0,$$

$$-\frac{1}{c}\frac{\partial H_z}{\partial t} - \frac{\partial E_y}{\partial x} + \frac{\partial E_x}{\partial y} = 0,$$

or
$$\operatorname{div} \mathbf{H} = 0, \text{ and } \operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t}.$$

Thus it appears that if we have any skew-symmetrical world tensor $\mathfrak{t}$ which is equal to the curl of a four-dimensional vector $\mathbf{P}$, and if we put

$$\mathfrak{t}_{yz} = H_x, \quad \mathfrak{t}_{zx} = H_y, \quad \mathfrak{t}_{xy} = H_z,$$
$$\mathfrak{t}_{\tau x} = E_x, \quad \mathfrak{t}_{\tau y} = E_y, \quad \mathfrak{t}_{\tau z} = E_z,$$

we shall have
$$\operatorname{div} \mathbf{H} = 0, \quad \operatorname{div} \mathbf{E} = V_t,$$
$$\operatorname{curl} \mathbf{H} = \mathbf{V}_1 + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t}, \quad \operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t},$$

where V_t is the time component of the vector divergence of the tensor $\mathfrak{t}$ and $\mathbf{V}_1$ is the component perpendicular to the time axis. These results of course are true for any set of axes x, y, z, τ in the four-dimensional world.

11. The Electromagnetic Equations.

Maxwell's equations of the electromagnetic field in the form given to them by H. A. Lorentz are (see Chapter I):

$$\operatorname{div} \mathbf{H} = 0, \quad \operatorname{div} \mathbf{E} = \rho,$$
$$\operatorname{curl} \mathbf{H} = \frac{1}{c}\left(\rho \mathbf{v} + \frac{\partial \mathbf{E}}{\partial t}\right), \quad \operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t},$$

where $\mathbf{H}$ is the magnetic field strength, $\mathbf{E}$ the electric field strength, ρ the density of electricity or charge per unit volume, and $\mathbf{v}$ the velocity of the electricity. Comparing these equations with those just obtained above, we see that if $V_t = \rho$ and $\mathbf{V}_1 = \rho\mathbf{v}/c$ the two sets of equations become identical.

If e is the charge in a volume S which is at rest relatively to the axes used, so that $\rho = e/S$, and if we change to axes in which the charge is moving with velocity v, the volume changes to S/β so that the density of charge becomes $e\beta/S$. Hence, if we put $\rho' = \rho/\beta$, ρ' has the same value for any set of axes. ρ' may be called the Minkowski density of charge.

The Minkowski current density may be defined as the product of

ρ' and the Minkowski velocity, so that it is equal to $\rho' ic$, and its components are $\rho'\beta v_x = \rho v_x$, ρv_y, ρv_z, and $ic\rho$. Its time component is therefore ρc, and its component perpendicular to the t axis is $\rho\mathbf{v}$. If then $V_t = \rho$ and $\mathbf{V}_\perp = \rho\mathbf{v}/c$, it appears that $\mathbf{V}$ is equal to the Minkowski current divided by c.

Maxwell's electromagnetic equations therefore show that the three components of the magnetic field and the three components of the electric field may be regarded as the six components of a skew-symmetrical world tensor, which is called the electromagnetic field tensor. This tensor is equal to the curl of a world vector, which may be called the Minkowski vector potential, and its vector divergence is equal to the Minkowski current divided by the velocity of light.

As an example, suppose that we have a charge e in a magnetic field $H = H_y$ along the y axis, and suppose that the charge is at rest relative to the axes. If now we change to axes in which the charge has a velocity v_x along the x axis, we are rotating the axes in the (x, τ) plane through an angle the cosine of which is β. Then $H_y = \mathfrak{t}_{z\tau}$, and so H_y transforms in the same way as the product of two vectors, one along the z axis and one along the x axis. The rotation in the (x, τ) plane leaves a vector along z unchanged in position, but a vector along x is changed into one inclined to the x axis in the (x, τ) plane, and its components are proportional to β along x and $v_x\beta/c$ along τ. Thus the magnetic field H_y transforms into a magnetic field βH_y along y and an electric field $E_z = -v_x\beta H_y/c$ along z, because $E_z = \mathfrak{t}_{tz}$. Thus when the charge e is moving with velocity v_x along the x axis in a magnetic field of strength βH_y along the y axis, there is a force on it along the z axis equal to $\beta H_y e v_x/c$. The field E_z and the force $\beta H_y e v_x/c$ together give no resultant force on the particle along the z axis.

We see in this way that a magnetic field in one set of axes becomes a magnetic field and an electric field in another set. Magnetic and electric fields are therefore relative quantities, but the electromagnetic field tensor $\mathfrak{t}$ is the same in any set of axes, since it is equal to the curl of the Minkowski vector potential, and the latter is a four-dimensional vector.

References

1. *The Theory of Relativity.* L. Silberstein, 1914.
2. *Introduction to Theoretical Physics.* A. Haas. Vol. II.
3. *The Principle of Relativity.* E. Cunningham.

CHAPTER XIX

General Relativity and Gravitation

1. Principle of Equivalence.

We have seen that, since a gravitational field gives equal accelerations to material particles of all kinds, it is impossible, by observations on the motion of such particles in a laboratory, to distinguish between accelerations due to a gravitational field and accelerations due to an acceleration of the laboratory.

Einstein put forward a generalization of this principle which is known as the *principle of equivalence*. According to the principle of equivalence, the effects due to a uniform gravitational field are precisely the same as those due to a uniform acceleration of the material frame of reference relatively to which the phenomena are observed.

The uniform acceleration is supposed to be equal and opposite to the acceleration which the gravitational field gives to a particle of any kind. It follows that if a frame of reference moves with an acceleration equal to, and in the same direction as, that due to the gravitational field, the effects due to the field will be equal and opposite to those due to the acceleration, and there will be no observable effects.

The principle of equivalence is supposed to be true for electrical and optical phenomena as well as for the motion of material particles.

According to this, if the gravitational fields of the sun and moon were uniform over the whole earth, there would be no observable effects on the earth due to these fields. The tides which are observed are attributed to the variation of the fields with the distance from the sun and moon.

A good illustration of the principle of equivalence is the following, due to Einstein. Imagine an observer working in a large completely closed box, and suppose he knows nothing of what goes on outside the box. Let the box be far from all other material bodies so that there is no gravitational field where it is. Now suppose a rope attached to the box and the rope pulled so that the box is made to move with an acceleration. The observer in the box would find that everything in the box when unsupported moved with an equal and opposite acceleration relative to the box. He would conclude probably that

the box was in a gravitational field. To keep a body at rest in the box would require a force proportional to the mass of the body. The observer could measure the acceleration by means of a pendulum, just as the acceleration of gravity is measured on the earth. If he could observe the path of a ray of light with sufficient accuracy, he would find it not exactly straight because of the acceleration of the box. He might conclude that the refractive index of the space in his box varied slightly from point to point

Suppose now that the box is supported in a uniform gravitational field so that it remains at rest. To the observer in the box, according to the principle of equivalence, everything will be the same as in the first case. It follows, for example, that light rays must be slightly deviated by a gravitational field.

The principle of equivalence does not mean that it is never possible to distinguish between gravitational fields and accelerations of the frame of reference. For example, the acceleration of gravity g observed on the earth cannot be attributed to an upward acceleration of the earth's surface equal to g, because such an acceleration would mean that the diameter of the earth was increasing all the time, which is of course impossible.

The principle of equivalence suggests that producing a gravitational field in a space is analogous to changing the frame of reference or co-ordinate system used to describe phenomena in the space. In the four-dimensional Minkowski world of the special theory of relativity the world lines of the particles represent phenomena independently of any co-ordinate system. Let us suppose that the system we are considering consists of particles subject only to gravitational forces—that is, suppose that there are no electric, magnetic, or other forces besides gravitational in the system. Einstein supposes that the special theory of relativity is true for such a system only when the masses of the particles are so small and the distances between them so large that the gravitational forces are negligible. In this case the world lines of the particles are all straight in any system of rectangular co-ordinates x, y, z, τ which may be adopted. The interval between two points A and B measured along a world line joining the points is equal to $\int_A^B ds$, and for a straight line between A and B $\delta\int_A^B ds = 0$, where δ denotes a variation from the straight line to any infinitely near line from A to B.

A line for which $\int_A^B ds$ is stationary, or for which $\delta\int_A^B ds = 0$, is called a *geodetic line* or *geodesic*.

2. Curvilinear Co-ordinates. World Lines and Geodesics.

The interval ds between two neighbouring points is the same in any system of rectangular co-ordinates, so that this is true also of $\int_A^B ds$.

Suppose that we use a system of curvilinear co-ordinates instead of the rectangular ones. Then, since ds has the same value in any co-ordinate system, $\int_A^B ds$ along a world line will still be stationary in the curvilinear system. The world line will still be given by $\delta\int_A^B ds = 0$, where δ denotes a variation due to changing from the world line between A and B to any other line between A and B which is very near to the world line.

Now according to the principle of equivalence the effects due to a gravitational field are the same as the effects due to changing the co-ordinate system in a suitable way, so that we may assume that in a gravitational field the world line of a particle will be given by $\delta\int_A^B ds = 0$.

If, however, we use rectangular co-ordinates in the presence of a gravitational field $\delta\int_A^B ds = 0$ will give straight lines whereas the world lines in a gravitational field will not be straight, because the gravitational forces will produce curvature of the lines. We conclude, therefore, that rectangular co-ordinates are impossible in a gravitational field. This means that the gravitational field modifies space in such a way that it is impossible to choose co-ordinates in which ds^2 is everywhere equal to $dx^2 + dy^2 + dz^2 + d\tau^2$.

Since a very short element of any curve may be regarded as straight, we see that it is possible to choose rectangular co-ordinates in which any particular element ds of a world line is given by $ds^2 = dx^2 + dy^2 + dz^2 + d\tau^2$, but the same rectangular co-ordinates cannot be used for the successive elements ds of the line.

To see what this means let us take a simple case from ordinary geometry. Consider two points A and B lying in a plane, and let ds be an element of any line from A to B. If $\delta\int_A^B ds = 0$, the line is as short as possible. Let x, y be the rectangular co-ordinates of a point on the plane, so that $ds^2 = dx^2 + dy^2$. Suppose we draw two sets of equidistant parallel straight lines on the plane, one set parallel to the x axis and the other to the y axis. Let these lines be given by $x = 0, 1, 2, 3, 4 \ldots$ and $y = 0, 1, 2, 3, 4 \ldots$, and let the lines be numbered with numbers equal to the values of x or y at each line.

Suppose now that the sets of lines are distorted in any way so that they become curved. If, for example, they were drawn on a thin

sheet of rubber, by stretching the rubber they could be changed into two sets of curves. The small equal squares between the lines would be changed into unequal parallelograms. We suppose the line between the two points A and B to be fixed and not moved in any way by the distortion of the two sets of lines

We may use the numbers on the curved lines as co-ordinates of a point on the plane if we suppose that the distortion did not move any of the lines out of the plane. Thus a point half-way between the two x lines numbered 5 and 6 and the two y lines numbered 9 and 10 would have co-ordinates 5·5 and 9·5. It is clear that any such change from rectangular to curvilinear co-ordinates in the plane makes no difference to $\int_A^B ds$ and that $\delta\int_A^B ds = 0$ is still the condition for the line joining AB to be as short as possible, because the length of ds is the same in whatever system of co-ordinates it is expressed. If the co-ordinates of A are x_1, y_1 in the rectangular system and those of B x_2, y_2, then $AB^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$, and the equation of the straight line AB is

$$y - y_1 = (y_2 - y_1)\frac{x - x_1}{x_2 - x_1},$$

or $y = ax + b$, where a and b are constants. In the curvilinear system the equation of the line AB will not be linear so that the line may be said to be curved relatively to the curvilinear co-ordinates if we regard a line given by a linear equation as straight. The line given by $\delta\int_A^B ds = 0$ will be straight in rectangular co-ordinates and curved in curvilinear co-ordinates, but in any co-ordinates it will be the shortest line between A and B.

Such a change to curvilinear co-ordinates in a plane is analogous to changing from rectangular co-ordinates x, y, z, τ to curvilinear co-ordinates in the Minkowski world when there is no gravitational field. In the curvilinear system particles move along curved paths as though acted on by a field of force, or in other words the use of a co-ordinate system which is accelerated produces an apparent field which is exactly like a gravitational field.

Suppose now that when the two sets of equidistant parallel lines in the plane are distorted the plane itself is also distorted into a curved surface. Then a line from A to B on the surface cannot be straight, but $\delta\int_A^B ds = 0$ will still give the shortest line from A to B on the surface. On a curved surface it is not in general possible to set up a system of rectangular co-ordinates in which $ds^2 = dx^2 + dy^2$ everywhere, but any small element of the surface can be regarded as plane, and a small local set of rectangular axes can be drawn on it in which $ds^2 = dx^2 + dy^2$.

According to Einstein's theory, the distortion in this case is analogous to the introduction of a gravitational field into the Minkowski world. Einstein supposes that the gravitational field distorts or curves space so that rectangular co-ordinates in which $ds^2 = dx^2 + dy^2 + dz^2 + d\tau^2$, everywhere, become impossible, but $\delta\int_A^B ds = 0$ still gives the position of the world line of a particle from A to B.

According to this theory, then, the world line of a particle between two points A and B is always given by $\delta\int_A^B ds = 0$.

The gravitational field is a curvature or distortion of the Minkowski four-dimensional world, and the world lines are geodesics in this distorted space. The equation $\delta\int_A^B ds = 0$ is purely geometrical, and so all particles of whatever mass or nature move in the same way in a given field. That is, the gravitational acceleration is the same for all kinds of matter, as is found experimentally to be the case.

As we have seen, along the path of a ray of light in the Minkowski world, when there is no gravitational field, we have $ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2$, where $icdt$ has been put for $d\tau$, and since in this case

$$v^2 = \frac{dx^2 + dy^2 + dz^2}{dt^2} = c^2,$$

we get $ds = 0$ for any element of the line. This result of course must be true in any system of co-ordinates and, according to the principle of equivalence, it must also be true in a gravitational field. The world line of a ray of light is therefore a geodesic of zero length.

To determine the path of a particle or ray of light in a gravitational field it is therefore necessary to find in what way space is distorted near material bodies, so that ds may be expressed in terms of any set of co-ordinates chosen and the equation of the world line between two points found in terms of the co-ordinates by means of the equation $\delta\int_A^B ds = 0$. The co-ordinates used may be any which are possible in the curved space just as, for example, any co-ordinates may be used to fix the positions of points on the surface of a sphere, provided they conform to the geometry of the spherical surface.

3. General Expression for ds^2. Einstein's Problem.

In the curved Minkowski world any very small region may be regarded as not distorted, and a small local system of rectangular co-ordinates may be supposed drawn in it. In these local co-ordinates $ds^2 = dx^2 + dy^2 + dz^2 + d\tau^2$. Now let x_1, x_2, x_3, x_4 be taken as the co-ordinates in a system which can be used throughout the space. Then in the small region of the co-ordinates x, y, z, τ these rectangular

co-ordinates must be functions of x_1, x_2, x_3, x_4. Hence we have

$$dx = \frac{\partial x}{\partial x_1}dx_1 + \frac{\partial x}{\partial x_2}dx_2 + \frac{\partial x}{\partial x_3}dx_3 + \frac{\partial x}{\partial x_4}dx_4,$$

with similar equations for dy, dz, and $d\tau$.

Substituting these values of dx, dy, dz, and $d\tau$ in the equation $ds^2 = dx^2 + dy^2 + dz^2 + d\tau^2$, we get

$$\begin{aligned} ds^2 = g_{11}dx_1^2 + g_{22}dx_2^2 + g_{33}dx_3^2 + g_{44}dx_4^2 \\ + 2g_{12}dx_1dx_2 + 2g_{13}dx_1dx_3 + 2g_{14}dx_1dx_4 \\ + 2g_{23}dx_2dx_3 + 2g_{24}dx_2dx_4 + 2g_{34}dx_3dx_4, \end{aligned}$$

where the g's are functions of the co-ordinates x_1, x_2, x_3, x_4. For example, $g_{11} = \left(\frac{\partial x}{\partial x_1}\right)^2 + \left(\frac{\partial y}{\partial x_1}\right)^2 + \left(\frac{\partial z}{\partial x_1}\right)^2 + \left(\frac{\partial \tau}{\partial x_1}\right)^2$, and x, y, z, τ are functions of x_1, x_2, x_3, x_4. This expression for ds^2 may be written $ds^2 = \Sigma g_{\mu\nu}dx_\mu dx_\nu$, in which it is understood that all the sixteen combinations of the four values 1, 2, 3, 4 of μ and ν are to be summed, and that $g_{\mu\nu} = g_{\nu\mu}$. An expression for ds^2 of this form holds of course in any system of co-ordinates.

The relations between the g's and the properties of the curved space had been previously worked out by Riemann and other mathematicians, so that Einstein was able to make use of their results. The problem is to find the relations between the g's which hold in any possible co-ordinate system in a gravitational field. The relations between the g's are sets of differential equations, and Einstein selected a set which seemed likely to be the correct set. In making this more or less arbitrary choice he was guided by the knowledge that Newton's law of gravitation is certainly a close approximation to the truth, so that it was necessary to choose equations which gave results differing little from Newton's theory. Also, in the absence of a gravitational field he supposed the four-dimensional space to be undistorted, so that the differential relations between the g's had to be such that they reduced to the relations for an undistorted space at great distances from matter.

Einstein succeeded in finding a set of equations satisfying these conditions, and he then worked out the theory of the motion of a small planet round a large attracting mass, and the path of a ray of light near the sun. The results obtained have been found to agree with the facts, and his theory is therefore regarded as probably correct.

4. Theory of Tensors.

Thus Einstein's theory depends on the differential geometry of four-dimensional space, which in turn depends on the theory of tensors. In a space of four dimensions let x_1, x_2, x_3, x_4 be the co-ordinates of a

point, and let the components of a small displacement be dx_1, dx_2, dx_3, dx_4. If now we change to any other set of co-ordinates x'_1, x'_2, x'_3, x'_4 the components of the small displacement will be given by

$$dx'_1 = \frac{\partial x'_1}{\partial x_1}dx_1 + \frac{\partial x'_1}{\partial x_2}dx_2 + \frac{\partial x'_1}{\partial x_3}dx_3 + \frac{\partial x'_1}{\partial x_4}dx_4,$$

with similar equations for dx'_2, dx'_3, and dx'_4. These equations may be written

$$dx'_\mu = \frac{\partial x'_\mu}{\partial x_\sigma}dx_\sigma,$$

where σ is to be given the values 1, 2, 3, 4, and the four terms are to be added.

In what follows it will be understood that, when a suffix occurs twice in a term, the term stands for the sum of its four values corresponding to the four values of the suffix 1, 2, 3, and 4.

A displacement is a vector, and any other vector which is transformed from one set of co-ordinates to another in the same way as a displacement is called a *contravariant vector*. A contravariant vector with components A^1, A^2, A^3, A^4 referred to x_1, x_2, x_3, x_4 transforms into one with components A'^1, A'^2, A'^3, A'^4 referred to x'_1, x'_2, x'_3, x'_4 where

$$A'^1 = \frac{\partial x'_1}{\partial x_1}A^1 + \frac{\partial x'_1}{\partial x_2}A^2 + \frac{\partial x'_1}{\partial x_3}A^3 + \frac{\partial x'_1}{\partial x_4}A^4,$$

with similar equations for A'^2, A'^3, and A'^4.

Thus a contravariant vector A^σ transforms into

$$A'^\mu = \frac{\partial x'_\mu}{\partial x_\sigma}A^\sigma,$$

where $\mu = 1, 2, 3,$ or 4 and $\sigma = 1, 2, 3,$ or 4. Since σ appears twice the term $\frac{\partial x'_\mu}{\partial x_\sigma}A^\sigma$ is understood to be summed.

If ϕ is a scalar function of position, the vector having components $\frac{\partial \phi}{\partial x_1}, \frac{\partial \phi}{\partial x_2}, \frac{\partial \phi}{\partial x_3}, \frac{\partial \phi}{\partial x_4}$ is transformed by the formula

$$\frac{\partial \phi}{\partial x'_1} = \frac{\partial x_1}{\partial x'_1}\frac{\partial \phi}{\partial x_1} + \frac{\partial x_2}{\partial x'_1}\frac{\partial \phi}{\partial x_2} + \frac{\partial x_3}{\partial x'_1}\frac{\partial \phi}{\partial x_3} + \frac{\partial x_4}{\partial x'_1}\frac{\partial \phi}{\partial x_4},$$

since we have $d\phi = \frac{\partial \phi}{\partial x_1}dx_1 + \frac{\partial \phi}{\partial x_2}dx_2 + \frac{\partial \phi}{\partial x_3}dx_3 + \frac{\partial \phi}{\partial x_4}dx_4$; with similar equations for $\frac{\partial \phi}{\partial x'_2}$, $\frac{\partial \phi}{\partial x'_3}$, and $\frac{\partial \phi}{\partial x'_4}$. A vector which transforms in this

way is called a *covariant vector*. If A_μ denotes a component of a covariant vector, then

$$A'_\mu = \frac{\partial x_\sigma}{\partial x'_\mu} A_\sigma.$$

Contravariant vectors are indicated by a raised suffix and covariant vectors by a lowered suffix. The contravariant vector dx_μ, however, is usually written with a lowered suffix.

Tensors are quantities having components which transform in the same way as the products of the components of two or more vectors. A tensor which transforms like the products of the components of two covariant vectors is written $A_{\mu\nu}$, and transforms thus:

$$A'_{\mu\nu} = \frac{\partial x_\sigma}{\partial x'_\mu} \frac{\partial x_\tau}{\partial x'_\nu} A_{\sigma\tau}.$$

In this expression σ and τ occur twice, so that there are sixteen terms to be summed. The covariant tensor $A_{\mu\nu}$ has sixteen components,

$$\begin{matrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44}. \end{matrix}$$

In the same way, we define contravariant tensors $A^{\mu\nu}$, which transform like the products of components of contravariant vectors, so that

$$A'^{\mu\nu} = \frac{\partial x'_\mu}{\partial x_\sigma} \frac{\partial x'_\nu}{\partial x_\tau} A^{\sigma\tau};$$

and mixed tensors A^ν_μ, for which

$$A'^\nu_\mu = \frac{\partial x_\sigma}{\partial x'_\mu} \frac{\partial x'_\nu}{\partial x_\tau} A^\tau_\sigma.$$

Thus any component of a tensor in co-ordinates x'_1, x'_2, x'_3, x'_4 is equal to the sum of a number of terms each of which is proportional to one of its components in any other co-ordinates x_1, x_2, x_3, x_4. It follows that if a tensor is zero in any co-ordinates, that is, if all its components are zero, it will be zero in any other co-ordinate system.

If then any law of nature expressing results, relatively to a material frame of reference, in terms of co-ordinates x_1, x_2, x_3, x_4 can be expressed by an equation $T = 0$, where T is a tensor, then in terms of any other co-ordinates x'_1, x'_2, x'_3, x'_4 the law will be expressed by $T' = 0$, where T' is the tensor into which T is transformed by the change from x_1, x_2, x_3, x_4 to x'_1, x'_2, x'_3, x'_4.

According to Einstein's general principle of relativity all the laws of nature can be expressed by equations of the form $T = 0$, so that all such laws, when so expressed, are independent of the co-ordinate

system used. In particular the law of gravitation must be capable of being put into the form $T = 0$ like any other law of nature.

Tensors having two suffixes are said to be of the second rank, and vectors may be called tensors of the first rank. Tensors which transform like the products of more than two vector components may be defined in a similar way.

The product of a component $A_{\mu\nu}$ of a tensor with a component $B_{\sigma\tau}$ of another tensor is a component of a tensor $(A_{\mu\nu}B_{\sigma\tau})$, the components of which are all the products of the components like $A_{\mu\nu}$ with those like $B_{\sigma\tau}$. The product of two vectors A_μ and B_ν is a tensor $A_\mu B_\nu = C_{\mu\nu}$. The product of $A_{\mu\nu}$ and B^σ_λ is $A_{\mu\nu}B^\sigma_\lambda = C^\sigma_{\mu\nu\lambda}$.

If we multiply A_μ by B^μ we get $A_\mu B^\mu$, which is equal to

$$A_1B^1 + A_2B^2 + A_3B^3 + A_4B^4,$$

and so is a scalar quantity having only one component or value. The product $A_\mu B^\mu$ is called the *inner product*, to distinguish it from the ordinary product $A_\mu B^\nu$, which is a tensor having sixteen components, viz.:

$$\begin{matrix} A_1B^1 & A_1B^2 & A_1B^3 & A_1B^4 \\ A_2B^1 & A_2B^2 & A_2B^3 & A_2B^4 \\ A_3B^1 & A_3B^2 & A_3B^3 & A_3B^4 \\ A_4B^1 & A_4B^2 & A_4B^3 & A_4B^4. \end{matrix}$$

If an upper and lower suffix of a mixed tensor are both denoted by the same letter then the summation rule makes the tensor equal to the sum of four tensors, thus, for example if in $A^\sigma_{\mu\nu}$ we put $\sigma = \nu$ we get

$$A^\nu_{\mu\nu} = A^1_{\mu 1} + A^2_{\mu 2} + A^3_{\mu 3} + A^4_{\mu 4}.$$

The sum may be denoted by A_μ, and so denotes a vector with components A_1, A_2, A_3, A_4. Thus putting $\sigma = \nu$ in $A^\sigma_{\mu\nu}$ changes it to a vector A_μ. This operation is called *contraction*. If we contract A^ν_μ we get

$$A^1_1 + A^2_2 + A^3_3 + A^4_4,$$

which is an *invariant* (that is, retains the same value in all systems of co-ordinates), because

$$A'^\mu_\mu = \frac{\partial x_\sigma}{\partial x'_\mu}\frac{\partial x'_\mu}{\partial x_\sigma}A^\sigma_a = A^\sigma_\sigma.$$

It is important to be able to find out whether any quantity is a tensor or not. This can be done by finding its equations of transformation, or by expressing it as the sum or product of quantities known to be tensors. A quantity which on inner multiplication by any contravariant or covariant vector always gives a tensor is a tensor.

Thus let the product $A_{\mu\nu} B^\nu$ be a covariant vector for any B^ν. Then

$$A'_{\mu\nu} B'^{\nu} = \frac{\partial x_\sigma}{\partial x'_\mu} A_{\sigma\tau} B^\tau.$$

Also
$$B^\tau = \frac{\partial x_\tau}{\partial x'_\nu} B'^\nu,$$

so that
$$B'^\nu \left(A'_{\mu\nu} - \frac{\partial x_\sigma}{\partial x'_\mu} \frac{\partial x_\tau}{\partial x'_\nu} A_{\sigma\tau}\right) = 0.$$

Since B'^ν is arbitrary, it follows that

$$A'_{\mu\nu} = \frac{\partial x_\sigma}{\partial x'_\mu} \frac{\partial x_\tau}{\partial x'_\nu} A_{\sigma\tau},$$

which shows that $A_{\mu\nu}$ is a tensor.

5. The Fundamental Tensor.

Consider the equation $ds^2 = g_{\mu\nu}\, dx_\mu\, dx_\nu$, where $g_{\mu\nu} = g_{\nu\mu}$. It can be shown at once by actual transformation that $g_{\mu\nu}$ is a covariant tensor. Thus, since

$$dx_\mu = \frac{\partial x_\mu}{\partial x'_\sigma} dx'_\sigma \quad \text{and} \quad dx_\nu = \frac{\partial x_\nu}{\partial x'_\tau} dx'_\tau,$$

we have
$$g_{\mu\nu}\, dx_\mu\, dx_\nu = g_{\mu\nu} \frac{\partial x_\mu}{\partial x'_\sigma} \frac{\partial x_\nu}{\partial x'_\tau} dx'_\sigma\, dx'_\tau.$$

If this is written in the form $g'_{\sigma\tau}\, dx'_\sigma\, dx'_\tau$, we therefore have

$$g'_{\sigma\tau} = \frac{\partial x_\mu}{\partial x'_\sigma} \frac{\partial x_\nu}{\partial x'_\tau} g_{\mu\nu},$$

which shows that $g_{\mu\nu}$ is a covariant tensor. It is called the fundamental tensor. The product $g_{\mu\nu}\, dx_\mu\, dx_\nu$ is thus of the form $A_{\mu\nu} B^\mu C^\nu$ and so is an inner product, and invariant, as it should be since ds is the same in any co-ordinates.

The determinant

$$\begin{vmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{vmatrix} = |\, g_{\mu\nu} \,|$$

is denoted by g. If we erase from this determinant the ith row and the kth column, and multiply by $(-1)^{i+k}$, we get a minor D_{ik} of the determinant. This minor divided by the determinant is denoted by g^{ik}, so that $g^{ik} = D_{ik}/g$. Thus, for example,

$$g^{23} = -\begin{vmatrix} g_{11} & g_{12} & g_{14} \\ g_{31} & g_{32} & g_{34} \\ g_{41} & g_{42} & g_{44} \end{vmatrix} \div |\, g_{\mu\nu} \,|.$$

From this it follows that $g^\nu_\sigma = g_{\mu\sigma}g^{\mu\nu} = 1$ when $\sigma = \nu$, because

$$g_{\mu\sigma}g^{\mu\nu} = g_{1\sigma}g^{1\nu} + g_{2\sigma}g^{2\nu} + g_{3\sigma}g^{3\nu} + g_{4\sigma}g^{4\nu}$$

and

$$g = \sum_{h=1}^{h=4} g_{h\nu}D_{h\nu} = \sum_{i=1}^{i=4} g_{\nu i}D_{\nu i}.$$

In the same way $g_{\mu\sigma}g^{\mu\nu} = 0$, i.e. $g^\nu_\sigma = 0$, when σ is not equal to ν, because $g_{\mu\sigma}D_{\mu\nu}$ gives a determinant with two columns identical, which is therefore equal to zero.

Hence, if A^ν is any contravariant vector,

$$g^\nu_\sigma A^\sigma = g^\nu_1 A^1 + g^\nu_2 A^2 + g^\nu_3 A^3 + g^\nu_4 A^4 = A^\nu,$$

since $g^\nu_\sigma = 0$ when $\sigma \neq \nu$, and $g^\nu_\sigma = 1$ when $\sigma = \nu$.

The three fundamental tensors are

$$g_{\mu\nu},\ g^\nu_\mu,\ g^{\mu\nu}.$$

From any covariant tensor $A_{\mu\nu}$ we can get a mixed tensor $A^\nu_\mu = g^{\nu\alpha}A_{\mu\alpha}$, a contravariant tensor $A^{\mu\nu} = g^{\mu\alpha}g^{\nu\beta}A_{\alpha\beta}$, and an invariant or scalar $A = g^{\mu\nu}A_{\mu\nu}$.

6. Equations of a Geodesic. Christoffel's Symbols.

The world line of a particle is a geodesic determined by $\delta\int_A^B ds = 0$, where δ indicates a variation from the line itself, between two points A and B on it, to any other line which is very close to the world line. The variations are zero at A and B.

We have $ds^2 = g_{\mu\nu}dx_\mu dx_\nu = g_{hk}dx_h dx_k$, so that

$$\delta\int_A^B ds = \delta\int\sqrt{g_{hk}dx_h dx_k} = 0.$$

Also

$$\delta\int\sqrt{g_{hk}dx_h dx_k} = \delta\int\frac{g_{hk}dx_h dx_k}{\sqrt{g_{hk}dx_h dx_k}} = \delta\int g_{hk}\frac{dx_h}{ds}\frac{dx_k}{ds}ds.$$

Now

$$\delta\left(g_{hk}\frac{dx_h}{ds}\frac{dx_k}{ds}\right) = \delta g_{hk}\frac{dx_h}{ds}\frac{dx_k}{ds} + g_{hk}\left\{\frac{dx_h}{ds}\delta\left(\frac{dx_k}{ds}\right) + \frac{dx_k}{ds}\delta\left(\frac{dx_h}{ds}\right)\right\}.$$

Also $\delta g_{hk} = \dfrac{\partial g_{hk}}{\partial x_p}\delta x_p$, and the last two terms in the above equation are equal (as is seen at once by simply interchanging h and k, which leaves the same summations as before), so that

$$\delta\left(g_{hk}\frac{dx_h}{ds}\frac{dx_k}{ds}\right) = \frac{\partial g_{hk}}{\partial x_p}\delta x_p\frac{dx_h}{ds}\frac{dx_k}{ds} + 2g_{hp}\frac{dx_h}{ds}\delta\left(\frac{dx_p}{ds}\right).$$

In the last term, p has been written for convenience instead of k, but this obviously makes no difference. But

$$\delta\left(\frac{dx_p}{ds}\right) = \frac{d}{ds}(\delta x_p),$$

so that on multiplying by ds and integrating between A and B we get, for the geodesics,

$$\int_A^B \frac{\partial g_{hk}}{\partial x_p}\frac{dx_h}{ds}\frac{dx_k}{ds}\,\delta x_p\,ds + 2\int_A^B g_{hp}\frac{dx_h}{ds}\frac{d}{ds}(\delta x_p)\,ds = 0.$$

Integrating the second integral by parts, and remembering that the variation δ is zero at A and B, we get

$$\int_A^B \left\{\frac{\partial g_{hk}}{\partial x_p}\frac{dx_h}{ds}\frac{dx_k}{ds} - 2\frac{d}{ds}\left(g_{hp}\frac{dx_h}{ds}\right)\right\}\delta x_p\,ds = 0.$$

Since the variations δx_p are arbitrary, it follows that

$$\frac{\partial g_{hk}}{\partial x_p}\frac{dx_h}{ds}\frac{dx_k}{ds} - 2\frac{d}{ds}\left(g_{hp}\frac{dx_h}{ds}\right) = 0,$$

or $$g_{hp}\frac{d^2x_h}{ds^2} + \frac{dg_{hp}}{ds}\frac{dx_h}{ds} - \frac{1}{2}\frac{dx_h}{ds}\frac{dx_k}{ds}\frac{\partial g_{hk}}{\partial x_p} = 0.$$

But $$\frac{dg_{hp}}{ds} = \frac{\partial g_{hp}}{\partial x_k}\frac{dx_k}{ds},$$

and $$\frac{\partial g_{hp}}{\partial x_k}\frac{dx_k}{ds}\frac{dx_h}{ds} = \frac{\partial g_{kp}}{\partial x_h}\frac{\partial x_h}{\partial s}\frac{\partial x_k}{\partial s},$$

so that, putting $$\begin{bmatrix} h\ k \\ p \end{bmatrix} = \frac{1}{2}\left(\frac{\partial g_{hp}}{\partial x_k} + \frac{\partial g_{kp}}{\partial x_h} - \frac{\partial g_{hk}}{\partial x_p}\right),$$

we get $$g_{hp}\frac{d^2x_h}{ds^2} + \begin{bmatrix} h\ k \\ p \end{bmatrix}\frac{dx_h}{ds}\frac{dx_k}{ds} = 0.$$

The symbol $\begin{bmatrix} h\ k \\ p \end{bmatrix}$ is called a *Christoffel's three-index symbol*. Another kind of three-index symbol is defined by

$$\begin{Bmatrix} h\ k \\ r \end{Bmatrix} = g^{rp}\begin{bmatrix} h\ k \\ p \end{bmatrix}.$$

If now we multiply the equation

$$g_{hp}\frac{d^2x_h}{ds^2} + \begin{bmatrix} h\ k \\ p \end{bmatrix}\frac{dx_h}{ds}\frac{dx_k}{ds} = 0$$

by g^{rp}, and sum for p in the usual way we get, since $g_r^h = g_{hp} g^{rp}$ is equal to unity when $r = h$ and zero when $r \neq h$,

$$\frac{d^2x_r}{ds^2} = -\begin{Bmatrix} h\ k \\ r \end{Bmatrix} \frac{dx_h}{ds} \frac{dx_k}{ds}.$$

There are four equations of this type, corresponding to the values 1, 2, 3, 4 of r. Since ds is an element of a world line, these equations may be said to be the equations of motion of a particle moving along the world line of which ds is an element. When the g's are constants, $\begin{Bmatrix} h\ k \\ r \end{Bmatrix} = 0$, so that $\dfrac{d^2x_r}{ds^2} = 0$, which indicates that the world lines are straight, or that the particle moves with uniform velocity in a straight line.

7. Covariant Differentiation.

Let ϕ be an invariant function of position so that, since ds is also invariant, $d\phi/ds$ is invariant. $d\phi/ds$ is a vector and its components are $\partial\phi/\partial x_1$, $\partial\phi/\partial x_2$, $\partial\phi/\partial x_3$, $\partial\phi/\partial x_4$. The components of course depend on the co-ordinates chosen, but $d\phi/ds$ is the same in any co-ordinates.

We have
$$\frac{d\phi}{ds} = \frac{\partial\phi}{\partial x_h}\frac{dx_h}{ds},$$

so that
$$\frac{d^2\phi}{ds^2} = \frac{d}{ds}\left(\frac{\partial\phi}{\partial x_h}\right)\frac{dx_h}{ds} + \frac{\partial\phi}{\partial x_h}\frac{d^2x_h}{ds^2},$$

$$\frac{d}{ds}\frac{\partial\phi}{\partial x_h} = \frac{\partial}{\partial x_k}\left(\frac{\partial\phi}{\partial x_h}\right)\frac{dx_k}{ds},$$

and therefore
$$\frac{d^2\phi}{ds^2} = \frac{\partial^2\phi}{\partial x_h \partial x_k}\frac{dx_h}{ds}\frac{dx_k}{ds} + \frac{\partial\phi}{\partial x_h}\frac{d^2x_h}{ds^2}.$$

In the last term, of course, h can be changed to r if desired. Using the equation

$$\frac{d^2x_r}{ds^2} = -\begin{Bmatrix} h\ k \\ r \end{Bmatrix} \frac{dx_h}{ds} \frac{dx_k}{ds},$$

we then get
$$\frac{d^2\phi}{ds^2} = \left[\frac{\partial^2\phi}{\partial x_h \partial x_k} - \begin{Bmatrix} h\ k \\ r \end{Bmatrix}\frac{\partial\phi}{\partial x_r}\right]\frac{dx_h}{ds}\frac{dx_k}{ds}.$$

Now $d^2\phi/ds^2$ is an invariant or tensor of zero rank, and dx_h and dx_k are arbitrary contravariant vectors, so that $\dfrac{\partial^2\phi}{\partial x_h \partial x_k} - \begin{Bmatrix} h\ k \\ r \end{Bmatrix}\dfrac{\partial\phi}{\partial x_r}$ must be a covariant tensor of the second rank (cf. the end of section 4). If

then we put $A_h = \partial\phi/\partial x_h$, and denote the tensor just obtained by A_{hk}, we get

$$A_{hk} = \frac{\partial A_h}{\partial x_k} - \begin{Bmatrix} h\ k \\ r \end{Bmatrix} A_r.$$

A_{hk} is called the *covariant derivative* of A_h with respect to x_k. When the g's are constants so that $\begin{Bmatrix} h\ k \\ r \end{Bmatrix} = 0$, then $A_{hk} = \dfrac{\partial A_h}{\partial x_k}$, as in ordinary differentiation.

If A_λ, B_μ are covariant vectors, and $A_{\lambda\nu}$, $B_{\mu\nu}$ their covariant derivatives, the above formula gives

$$A_{\lambda\nu} B_\mu + A_\lambda B_{\mu\nu} = \frac{\partial(A_\lambda B_\mu)}{\partial x_\nu} - \begin{Bmatrix} \lambda\ \nu \\ \epsilon \end{Bmatrix} A_\epsilon B_\mu - \begin{Bmatrix} \mu\ \nu \\ \epsilon \end{Bmatrix} A_\lambda B_\epsilon.$$

Putting $A_{\lambda\mu\nu} = A_{\lambda\nu} B_\mu + A_\lambda B_{\mu\nu}$, $A_{\epsilon\mu} = A_\epsilon B_\mu$, and $A_{\lambda\epsilon} = A_\lambda B_\epsilon$,

we get
$$A_{\lambda\mu\nu} = \frac{\partial A_{\lambda\mu}}{\partial x_\nu} - \begin{Bmatrix} \lambda\ \nu \\ \epsilon \end{Bmatrix} A_{\epsilon\mu} - \begin{Bmatrix} \mu\ \nu \\ \epsilon \end{Bmatrix} A_{\lambda\epsilon}.$$

This gives the covariant derivative with respect to x_ν of the tensor $A_{\lambda\mu}$.

8. The Riemann Tensor.

Now obtain the second covariant derivative of a vector A_μ with respect first to x_ν, then to x_σ. Denoting it by $A_{\mu\nu\sigma}$, we get

$$A_{\mu\nu\sigma} = \frac{\partial}{\partial x_\sigma}\left[\frac{\partial A_\mu}{\partial x_\nu} - \begin{Bmatrix} \mu\ \nu \\ \rho \end{Bmatrix} A_\rho\right] - \begin{Bmatrix} \mu\ \sigma \\ \epsilon \end{Bmatrix}\left[\frac{\partial A_\epsilon}{\partial x_\nu} - \begin{Bmatrix} \epsilon\ \nu \\ \rho \end{Bmatrix} A_\rho\right] - \begin{Bmatrix} \nu\ \sigma \\ \epsilon \end{Bmatrix}\left[\frac{\partial A_\mu}{\partial x_\epsilon} - \begin{Bmatrix} \mu\ \epsilon \\ \rho \end{Bmatrix} A_\rho\right].$$

Reversing the order of the differentiation, we get a similar expression for $A_{\mu\sigma\nu}$, and so obtain

$$A_{\mu\nu\sigma} - A_{\mu\sigma\nu} = A_\rho\left[\begin{Bmatrix} \mu\ \sigma \\ \epsilon \end{Bmatrix}\begin{Bmatrix} \epsilon\ \nu \\ \rho \end{Bmatrix} - \begin{Bmatrix} \mu\ \nu \\ \epsilon \end{Bmatrix}\begin{Bmatrix} \epsilon\ \sigma \\ \rho \end{Bmatrix} + \frac{\partial}{\partial x_\nu}\begin{Bmatrix} \mu\ \sigma \\ \rho \end{Bmatrix} - \frac{\partial}{\partial x_\sigma}\begin{Bmatrix} \mu\ \nu \\ \rho \end{Bmatrix}\right].$$

The difference $A_{\mu\nu\sigma} - A_{\mu\sigma\nu}$ is a tensor, and A_ρ is an arbitrary vector so that we see that the expression in the bracket must be a tensor. It is called the Riemann tensor, and is denoted by $R^\rho_{\mu\nu\sigma}$, since when multiplied by A_ρ it gives $A_{\mu\nu\sigma} - A_{\mu\sigma\nu}$.

If the g's are constants, then all the three index symbols vanish, so that the Riemann tensor is then zero. The equations $R^\rho_{\mu\nu\sigma} = 0$ are a set of differential equations between the g's, and since $R^\rho_{\mu\nu\sigma}$ is a tensor it follows that if the equations $R^\rho_{\mu\nu\sigma} = 0$ are true in any one set of co-ordinates then they are true in any other set into which the

first set can be transformed. When the g's are constants there is no gravitational field, so that the equations $R^{\rho}_{\mu\nu\sigma} = 0$ are the relations between the g's which hold when there is no gravitational field, or when the space is undistorted, or, as we may say, is flat. Of course, if we use curvilinear co-ordinates in flat space, particles will move along curved lines as though acted on by gravitational forces, but these apparent forces are due to the acceleration of the co-ordinates and not to gravitation. If $R^{\rho}_{\mu\nu\sigma} = 0$ then rectangular axes for which $ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2$ are possible, since the space is flat.

9. Einstein's Law of Gravitation.

We are now in a position to consider Einstein's law of gravitation, or the relations between the g's which hold in curved space. The g's are analogous to the gravitational potential of Newton's theory. To see this, consider an element ds of a world line and suppose that its components dx_1, dx_2, dx_3, dx_4 are all zero except dx_1, so that $ds^2 = g_{11}dx_1^2$. Then, if $dx_4 = icdt$, we have $\left(\frac{ds}{dt}\right)^2 = -g_{11}c^2$. Now if ϕ is the gravitational potential in Newton's theory, $\frac{1}{2}v^2 = A - \phi$, where v is the velocity of a particle of unit mass, and A is a constant. Hence, putting $v = ds/dt$, we get $\left(\frac{ds}{dt}\right)^2 = 2(A - \phi)$. Thus we see that g_{11} is analogous to ϕ. In Newton's theory in empty space we have

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = 0,$$

so that, since Newton's theory is certainly a good approximation to the truth, we should expect the relations between the g's in empty space to be differential equations analogous to $\Delta\phi = 0$.

Also the relations must be consistent with $R^{\rho}_{\mu\nu\sigma} = 0$, but they must be less stringent, so as to admit gravitational fields or distortion of space.

It occurred to Einstein to try the contracted Riemann tensor $R_{\mu\nu} = 0$, and this guess has proved satisfactory. $R_{\mu\nu} = 0$ is a set of differential equations between the g's, of the second order like $\Delta\phi = 0$, and clearly consistent with $R^{\rho}_{\mu\nu\sigma} = 0$. The equations are tensor equations, and therefore consistent with the general principle of relativity. No other tensor has been discovered which can be used instead of $R_{\mu\nu}$.

Putting $\sigma = \rho$ in the expression for $R^{\rho}_{\mu\nu\sigma}$, we get

$$R_{\mu\nu} = \frac{\partial}{\partial x_\nu}\begin{Bmatrix}\mu\rho\\ \rho\end{Bmatrix} - \frac{\partial}{\partial x_\rho}\begin{Bmatrix}\mu\nu\\ \rho\end{Bmatrix} + \begin{Bmatrix}\mu\rho\\ \epsilon\end{Bmatrix}\begin{Bmatrix}\nu\epsilon\\ \rho\end{Bmatrix} - \begin{Bmatrix}\mu\nu\\ \epsilon\end{Bmatrix}\begin{Bmatrix}\epsilon\rho\\ \rho\end{Bmatrix} = 0.$$

According to Einstein, these are the relations between the g's which must hold in any case whether the space is curved or flat. In flat space the more stringent condition $R^{\rho}_{\mu\nu\sigma} = 0$ is true.

The principle of equivalence can now be stated more precisely. The relations $R^{\rho}_{\mu\nu\sigma}=0$ and $R_{\mu\nu}=0$ are relations between the *second* differential coefficients of the g's. It follows that all laws which hold in flat space and depend on the g's and their *first* derivatives only will also hold in curved space. Laws which involve the second derivatives of the g's need not hold in both flat space and curved space.

10. The Field due to a Particle.

Now consider the gravitational field in the space surrounding a single heavy particle. The problem is to find a set of values of the g's which satisfy $R_{\mu\nu}=0$ and the special conditions of the case.

In flat space we could suppose the particle at the origin and use polar co-ordinates r, θ, ϕ, t; that is, let $x_1=r$, $x_2=\theta$, $x_3=\phi$, $x_4=ict$, so that $ds^2=dr^2+r^2d\theta^2+r^2\sin^2\theta d\phi^2-c^2dt^2$. In this case, $g_{11}=1$, $g_{22}=r^2$, $g_{33}=r^2\sin^2\theta$, $g_{44}=-c^2$, and $g_{\sigma\tau}=0$ when $\sigma\neq\tau$. These values, however, would satisfy $R^{\rho}_{\mu\nu\sigma}=0$, and not merely $R_{\mu\nu}=0$. Let us put $c=1$, and assume (changing for convenience the sign of ds^2) $g_{11}=-\epsilon^{\lambda}$, $g_{22}=-r^2$, $g_{33}=-r^2\sin^2\theta$, $g_{44}=\epsilon^{\nu}$, and $g_{\sigma\tau}=0$ when $\sigma\neq\tau$. If these are possible values of the g's, they will satisfy the equations $R_{\mu\nu}=0$. It is assumed that λ and ν are functions of r only.

With these g's it is easy to calculate the three index symbols. The determinant g is equal to $-\epsilon^{\lambda+\nu}r^4\sin^2\theta$, and $g^{\tau\sigma}=\dfrac{1}{g_{\sigma\sigma}}$, hence we get

$$\begin{Bmatrix}\sigma\tau\\ \alpha\end{Bmatrix}=\frac{1}{2g_{\alpha\alpha}}\left\{\frac{\partial g_{\alpha\sigma}}{\partial x_\tau}+\frac{\partial g_{\alpha\tau}}{\partial x_\sigma}-\frac{\partial g_{\sigma\tau}}{\partial x_\alpha}\right\},$$

which is not to be summed.

Using this expression we find the following values of $\begin{Bmatrix}\sigma\tau\\ \alpha\end{Bmatrix}$:

$$\begin{Bmatrix}11\\ 1\end{Bmatrix}=\tfrac{1}{2}\frac{d\lambda}{dr},\qquad \begin{Bmatrix}13\\ 3\end{Bmatrix}=\frac{1}{r},$$

$$\begin{Bmatrix}12\\ 2\end{Bmatrix}=\frac{1}{r},\qquad \begin{Bmatrix}14\\ 4\end{Bmatrix}=\tfrac{1}{2}\frac{d\nu}{dr},$$

$$\begin{Bmatrix}22\\ 1\end{Bmatrix}=-r\epsilon^{-\lambda},\qquad \begin{Bmatrix}23\\ 3\end{Bmatrix}=\cot\theta,$$

$$\begin{Bmatrix}33\\ 1\end{Bmatrix}=-r\sin^2\theta\epsilon^{-\lambda},\qquad \begin{Bmatrix}33\\ 2\end{Bmatrix}=-\sin\theta\cos\theta,$$

$$\begin{Bmatrix}44\\ 1\end{Bmatrix}=\tfrac{1}{2}\epsilon^{\nu-\lambda}\frac{d\nu}{dr}.$$

The rest of the three-index symbols are zero. Using these values of

$\left\{\begin{matrix}\sigma\tau\\ \alpha\end{matrix}\right\}$ we now find the values of $R_{\mu\nu}$. It is found that they are all zero except R_{11}, R_{22}, R_{33}, and R_{44}.

The values of these are found to be

$$R_{11} = \tfrac{1}{2}\frac{d^2\nu}{dr^2} - \tfrac{1}{4}\frac{d\lambda}{dr}\frac{d\nu}{dr} + \tfrac{1}{4}\left(\frac{d\nu}{dr}\right)^2 - \frac{1}{r}\frac{d\lambda}{dr},$$

$$R_{22} = e^{-\lambda}\left\{1 + \tfrac{1}{2}r\left(\frac{d\nu}{dr} - \frac{d\lambda}{dr}\right)\right\} - 1,$$

$$R_{33} = \sin^2\theta \cdot e^{-\lambda}\left\{1 + \tfrac{1}{2}r\left(\frac{d\nu}{dr} - \frac{d\lambda}{dr}\right)\right\} - \sin^2\theta,$$

$$R_{44} = e^{\nu-\lambda}\left\{-\tfrac{1}{2}\frac{d^2\nu}{dr^2} + \tfrac{1}{4}\frac{d\lambda}{dr}\frac{d\nu}{dr} - \tfrac{1}{4}\left(\frac{d\nu}{dr}\right)^2 - \frac{1}{r}\frac{d\nu}{dr}\right\}.$$

Since each of these is to vanish, we obtain four equations. The first and last equations give $\frac{\partial\lambda}{\partial r} = -\frac{\partial\nu}{\partial r}$; also, when $r = \infty$, λ and ν must both be zero; therefore $\lambda = -\nu$.

The other two equations then give

$$e^{\nu}\left(1 + r\frac{\partial\nu}{\partial r}\right) = 1,$$

so that $e^{\nu} = 1 - \frac{2m}{r}$, where m is an integration constant. This solution satisfies all the equations.

Thus it appears that the forms assumed for the g's can satisfy the equations $R_{\mu\nu} = 0$, and they give a space symmetrical about the origin, so that they are a possible set for the case of a heavy particle at the origin. The result obtained is

$$ds^2 = -\left(1 - \frac{2m}{r}\right)^{-1}dr^2 - r^2d\theta^2 - r^2\sin^2\theta\, d\phi^2 + \left(1 - \frac{2m}{r}\right)dt^2.$$

11. Planetary Orbit.

Now suppose that an infinitesimal particle is moving in the space near the heavy particle at the origin. To find its motion we may either use $\delta\int ds = 0$ or the equations of motion

$$\frac{d^2x_r}{ds^2} = -\left\{\begin{matrix}h\,k\\ r\end{matrix}\right\}\frac{dx_h}{ds}\frac{dx_k}{ds}.$$

Substituting the values found above for $\left\{\begin{matrix}h\,k\\ r\end{matrix}\right\}$ we get, when $r = 2$ so that $x_r = \theta$,

$$\frac{d^2\theta}{ds^2} = \cos\theta\,\sin\theta\left(\frac{d\phi}{ds}\right)^2 - \frac{2}{r}\frac{dr}{ds}\frac{d\theta}{ds}.$$

If initially the particle is moving in the plane $\theta = \pi/2$, then $d\theta/ds = 0$, and $\cos\theta = 0$, so that initially

$$\frac{d^2\theta}{ds^2} = 0,$$

The particle therefore remains in the plane $\theta = \pi/2$. With $\theta = \pi/2$, the equations for x_3 and x_4 are

$$\frac{d^2\phi}{ds^2} + \frac{2}{r}\frac{dr}{ds}\frac{d\phi}{ds} = 0,$$

$$\frac{d^2t}{ds^2} + \frac{d\nu}{dr}\frac{dr}{ds}\frac{dt}{ds} = 0.$$

Integrating these equations we get $r^2\dfrac{d\phi}{ds} = h$, and $\dfrac{dt}{ds} = \dfrac{a}{\gamma}$, where h and a are constants, and $\gamma = e^\nu = 1 - \dfrac{2m}{r}$. The equation for ds^2, with $d\theta = 0$ and $\sin\theta = 1$, gives $ds^2 = -\gamma^{-1}dr^2 - r^2 d\phi^2 + \gamma dt^2$. This with $\dfrac{dt}{ds} = \dfrac{a}{\gamma}$ becomes

$$1 = -\gamma^{-1}\left(\frac{dr}{ds}\right)^2 - r^2\left(\frac{d\phi}{ds}\right)^2 + a^2\gamma^{-1}.$$

Since $\dfrac{dr}{ds} = \dfrac{dr}{d\phi}\dfrac{d\phi}{ds}$, $\dfrac{d\phi}{ds} = \dfrac{h}{r^2}$, and $\gamma = 1 - \dfrac{2m}{r}$, we get

$$\left(\frac{dr}{d\phi}\right)^2\frac{h^2}{r^4} = a^2 - 1 + \frac{2m}{r} - \frac{h^2}{r^2} + \frac{2mh^2}{r^3}.$$

Putting $u = r^{-1}$, this becomes

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = \frac{a^2-1}{h^2} + \frac{2mu}{h^2} + 2mu^3.$$

The corresponding equation for Newton's law of gravitation is

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = \frac{a^2-1}{h^2} + \frac{2mu}{h^2},$$

where m is the mass of the particle at the origin, and $h = \dfrac{r^2 d\phi}{dt}$. We may conclude that m is the mass of the particle in Einstein's theory. The additional term $2mu^3$ is very small in the case of the planets of the solar system, and makes no appreciable difference except in the case of the planet Mercury.

On differentiating Einstein's equation with respect to ϕ and dividing by $2du/d\phi$, we find

$$\frac{d^2u}{d\phi^2} + u = \frac{m}{h^2} + 3mu^2.$$

The corresponding Newtonian equation is

$$\frac{d^2u}{d\phi^2} + u = \frac{m}{h^2},$$

the solution of which is

$$u = \frac{m}{h^2}(1 + e\cos\phi),$$

representing (if $e < 1$) an ellipse with the attracting particle at one focus, and perihelion at $\phi = 0$. The additional term $3mu^2$ in Einstein's equation is very small in comparison with m/h^2 in planetary orbits. We shall show that it corresponds to a slow rotation of the major axis of the ellipse.

Following the method of successive approximation, we substitute the approximate value of u as derived from the Newtonian equation, in Einstein's equation, and so find

$$\frac{d^2u}{d\phi^2} + u = \frac{m}{h^2} + \frac{3m^3}{h^4}(1 + \tfrac{1}{2}e^2 + 2e\cos\phi + \tfrac{1}{2}e^2\cos 2\phi),$$

the solution of which is

$$u = \left\{\frac{m}{h^2} + \frac{3m^3}{h^4}(1 + \tfrac{1}{2}e^2)\right\}(1 + e\cos\phi) + \frac{3m^3e}{h^4}(\phi\sin\phi - \tfrac{1}{6}e\cos 2\phi).$$

At perihelion u is a maximum, and $\dfrac{du}{d\phi} = 0$. Hence

$$0 = -\left\{\frac{m}{h^2} + \frac{3m^3}{h^4}(1 + \tfrac{1}{2}e^2)\right\}e\sin\phi + \frac{3m^3e}{h^4}(\sin\phi + \phi\cos\phi + \tfrac{1}{3}e\sin 2\phi).$$

If it were not for the term in $\phi\cos\phi$, this would give a perihelion value $\phi = 2\pi$ as in the Newtonian case. Actually the value is $\phi = 2\pi + \epsilon$, where ϵ is small. To find ϵ, substitute $2\pi + \epsilon$ for ϕ in the preceding equation. Thus

$$0 = -\frac{me}{h^2}\epsilon + \frac{3m^3e}{h^4}(2\pi),$$

neglecting terms of higher order; so that

$$\epsilon = \frac{6\pi m^2}{h^2}.$$

In the case of the planet Mercury this formula gives for the rotation of the major axis of the orbit in one hundred years approximately 43 sec. of arc.

It has been known to astronomers for a long time that, after allowing for all disturbing influences due to the other planets, there is a rotation of the perihelion of Mercury's orbit of about 43 sec. per

century which cannot be explained by the Newtonian theory. The fact that Einstein's theory gives just this observed rotation is a remarkable success for his theory. For the other planets the correction is inappreciable.

12. Deflection of Light Rays by the Sun.

The path of a ray of light in a gravitational field according to Einstein's theory is the same as that of a particle moving with the velocity of light. For such a particle $ds = 0$ where ds is any element of its path in the four-dimensional world.

For any particle we have, as already shown,

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = \frac{a^2 - 1}{h^2} + \frac{2mu}{h^2} + 2mu^3,$$

and

$$\frac{d^2u}{d\phi^2} + u = \frac{m}{h^2} + 3mu^2,$$

where $h = r^2\, d\phi/ds$. For a light ray, $ds = 0$ and therefore $h = \infty$. The differential equation becomes

$$\frac{d^2u}{d\phi^2} + u = 3mu^2.$$

When $m = 0$, this is satisfied by

$$u = \frac{1}{p}\cos\phi,$$

the equation of a straight line, p being the perpendicular from the origin.

Substitute this first approximation as before in the small term on the right. Thus

$$\frac{d^2u}{d\phi^2} + u = \frac{3m}{p^2}\,\frac{1 + \cos 2\phi}{2},$$

a solution of which is

$$u = \frac{1}{p}\cos\phi + \frac{3m}{p^2}\left(\tfrac{1}{2} - \tfrac{1}{6}\cos 2\phi\right),$$

where, as in the first approximation, the angle ϕ is measured from the point where u is a maximum. This equation will give $u = 0$ (or $r = \infty$) when $\phi = \frac{\pi}{2} + \epsilon$ or $\phi = -\frac{\pi}{2} - \epsilon$, where ϵ is small. To find ϵ, we have

$$0 = \frac{1}{p}(-\epsilon) + \frac{3m}{p^2}\left(\tfrac{1}{2} + \tfrac{1}{6}\right),$$

or

$$\epsilon = \frac{2m}{p}.$$

The asymptotic directions of the ray therefore deviate by $\frac{2m}{p}$ radians from the mean direction.

The final deviation of a ray which starts from an infinite distance on one side and travels to an infinite distance on the other side is therefore given by

$$\Delta = \frac{4m}{p}.$$

Thus the light from a star passing by the sun should be deviated at the earth through an angle $4m/p$.

We have taken the gravitational potential to be equal to m/r and the velocity of light to be unity so that our unit of mass is very large.

The gravitational attraction in dynes between two particles of masses m and m' gm. at a distance r cm. apart is Gmm'/r^2 where $G = 6{\cdot}66 \times 10^{-8}$. Hence if g is the acceleration of a particle at a distance r from a mass m then $g = Gm/r^2$. If the unit of mass is to be a mass which gives an acceleration equal to the velocity of light in centimetres per second to a particle at a distance from it equal to the number of centimetres light travels in 1 sec. then we have $3\times 10^{10} = Gm/9\times 10^{20}$, where m is this new unit of mass expressed in grammes. Hence $m = 27 \times 10^{30} \times 10^{8}/6{\cdot}66$, which is equal to 4×10^{38} gm.

The mass of the sun is nearly 2×10^{33} gm. or $\frac{2 \times 10^{33}}{4 \times 10^{38}} = 5 \times 10^{-6}$ in the new unit, and its radius is 7×10^{10} cm. or $2{\cdot}33$ when the unit of length is 3×10^{10} cm. Hence for a ray of light which just grazes the sun's surface the deviation Δ should be

$$\frac{4 \times 5 \times 10^{-6}}{2{\cdot}33} = 8{\cdot}6 \times 10^{-6}$$

or $\Delta = 1{\cdot}77$ sec. of arc.

The apparent positions of several stars when near the sun have been determined by photographing them during a total solar eclipse, and it has been found that there is a deviation of the light approximately equal to that given by Einstein's formula $\Delta = 4m/p$.

13. Displacement of Spectral Lines.

In the four-dimensional world the interval ds between two events is supposed to be the same in any co-ordinate system. Also all atoms of the same kind are supposed to be identical in structure, so that the interval ds between the emission of two successive light waves by an atom should be the same for all such atoms. The presence of a gravitational field should not alter ds because, as we have seen, the effects due to gravitation are the same as those due to changing co-ordinates.

Consider then two identical atoms, one at the surface of the sun and one on the earth. Let them both be at rest so that $dx = dy = dz = 0$. Then if ds is the interval between the emission of two light waves by either atom we have

$$ds^2 = \left(1 - \frac{2m}{r}\right) dt'^2$$

for the atom on the sun, and $ds^2 = dt^2$ for the atom on the earth, since $2m/r$ is negligible for the earth. Hence

$$\frac{dt}{dt'} = \sqrt{1 - \frac{2m}{r}} = 1 - \frac{m}{r}.$$

Thus dt' is slightly longer than dt, so that the spectral lines emitted by elements in the sun should be shifted slightly towards the red end of the spectrum m/r is equal to $2{\cdot}1 \times 10^{-6}$ so that the change of wave-length is very small, only a few thousandths of an Angstrom unit. As the result of very careful measurements mainly by C. E St. John it is now believed that this minute effect actually exists, in agreement with Einstein's theory.

REFERENCES

1. *General Theory of Relativity and Gravitation* L Silberstein.
2. *Space, Time, and Gravitation* A S. Eddington
3 *Mathematical Theory of Relativity* A. S Eddington.
4 *Space, Time, Matter.* H. Weyl.
5 *The Absolute Differential Calculus.* T. Levi-Civita.

CHAPTER XX

Mathematical Notes

1. Vector Analysis.

Many physical quantities are *vectors*, that is, quantities which have direction as well as magnitude, for example, velocity, acceleration, force, momentum, angular velocity. Quantities which are determined by a number alone without any reference to direction are called *scalars*, for example, mass, energy, electric charge. Two vectors are equal if they have equal numerical magnitudes and are in the same direction. Vectors can be represented by straight lines. The length of the line representing a vector is taken equal to the numerical magnitude of the vector, and the line is drawn parallel to the direction of the vector. It is necessary to regard the line as drawn *from* a point A *to* another point B, and the direction of the vector represented is from A to B, or an arrow head may be put on the line to indicate the direction of the vector along the line, as $\overrightarrow{AB}$. The vector $\mathbf{V}$ (in modern works usually printed thus in heavy type) represented by a line from A to B is sometimes said to be of opposite *sense* to the vector $\mathbf{V}'$ represented by the line from B to A. The word "sense", however, seems inappropriate, and we shall say that two such vectors are of opposite *sign*, so that $\mathbf{V} = -\mathbf{V}'$.

If $\mathbf{A}$ and $\mathbf{B}$ are two vectors, and $\mathbf{A} = \mathbf{B}$, then this vector equation means that the magnitudes of $\mathbf{A}$ and $\mathbf{B}$ are equal, and also that their directions are the same, so that they could both be represented by the same line drawn from one point to another.

If a vector $\mathbf{V}$ is represented by a line AB and another vector $\mathbf{U}$ by a line BC, then the vector $\mathbf{R}$ represented by the line AC is called the sum or resultant of the vectors $\mathbf{V}$ and $\mathbf{U}$, and this is expressed by the vector equation

$$\mathbf{R} = \mathbf{V} + \mathbf{U}.$$

The difference between $\mathbf{V}$ and $\mathbf{U}$ is taken to be the sum of $\mathbf{V}$ and $-\mathbf{U}$, got by drawing a line AB to represent $\mathbf{V}$ and a line BD equal but in the opposite direction to BC which represents $\mathbf{U}$, and taking AD to represent the difference $\mathbf{V} - \mathbf{U}$.

Relative to rectangular axes x, y, z any vector **V** has three components which may be denoted by V_x, V_y, V_z. These components are scalar quantities. A vector of magnitude V_x in the direction of the x axis may be represented by the symbol $V_x\mathbf{i}$, where **i** is a unit vector in the direction Ox. But it is sometimes convenient for brevity to denote the vector $V_x\mathbf{i}$ simply by $\mathbf{V}_x$. If l, m, n denote the cosines of the angles between a vector **V** and the axes x, y, z, then $V_x = \bar{V}l$, $V_y = \bar{V}m$, $V_z = \bar{V}n$, where $\bar{V}$ denotes the numerical magnitude of the vector **V**. If the vector **V** is represented by a line OP drawn from the origin O, then we can go from O to P by moving along the x axis a distance V_x, then parallel to the y axis a distance V_y, and finally parallel to the z axis a distance V_z. Thus we see that V is given by the vector equation

$$\mathbf{V} = \mathbf{V}_x + \mathbf{V}_y + \mathbf{V}_z.$$

If two vectors A and B are equal, so that $\mathbf{A} = \mathbf{B}$, it follows that $A_x = B_x$, $A_y = B_y$, and $A_z = B_z$. Any vector equation is therefore equivalent to three equations between the x, y, z components of the vectors. For example, if $\mathbf{R} = \mathbf{V} - \mathbf{U}$ then $\mathbf{R}_x = V_x - U_x$, $R_y = V_y - U_y$, $R_z = V_z - U_z$. The magnitude $\bar{V}$ of **V** is given by

$$\bar{V}^2 = V_x^2 + V_y^2 + V_z^2.$$

This equation is often written $V^2 = V_x^2 + V_y^2 + V_z^2$, because it is clear that only the magnitude of **V** can be involved, so that the bar above the V may be omitted. The magnitude of a vector **V** is often written $|\,\mathbf{V}\,|$ instead of $\bar{V}$. The equation

$$V^2 = V_x^2 + V_y^2 + V_z^2$$

with $V_x = \bar{V}l$, $V_y = \bar{V}m$, and $V_z = \bar{V}n$ gives the well-known result

$$1 = l^2 + m^2 + n^2.$$

If l, m, n are the direction cosines of **V**, as above, and λ, μ, ν the direction cosines of another vector **U**, then the angle θ between the two vectors is given by

$$\cos\theta = l\lambda + m\mu + n\nu,$$

for $\mathbf{V} = \mathbf{V}_x + \mathbf{V}_y + \mathbf{V}_z$, so that the component of **V** along **U** is equal to the sum of the components of $\mathbf{V}_x$, $\mathbf{V}_y$, and $\mathbf{V}_z$ along **U**, or $\lambda V_x + \mu V_y + \nu V_z$, which is equal to $l\lambda V + m\mu V + n\nu V$; and the component of **V** along **U** is equal to $V\cos\theta$.

2. Scalar Product and Vector Product of Two Vectors.

The product of the magnitude of the component of a vector $\mathbf{V}$ along another one $\mathbf{U}$ and the magnitude of $\mathbf{U}$ is called the *scalar product*, or *inner product*, of $\mathbf{V}$ and $\mathbf{U}$. It may be denoted by $(\mathbf{V}\cdot\mathbf{U})$. Thus $(\mathbf{V}\cdot\mathbf{U}) = VU\cos\theta$, where θ is the angle between $\mathbf{U}$ and $\mathbf{V}$. Also since $V\cos\theta = \lambda V_x + \mu V_y + \nu V_z$, and $U_x = \lambda U$, $U_y = \mu U$, $U_z = \nu U$, we have

$$(\mathbf{V}\cdot\mathbf{U}) = V_xU_x + V_yU_y + V_zU_z. \quad . \quad . \quad . \quad . \quad (1)$$

As an example of a scalar product we may consider the work done by a force $\mathbf{F}$ when its point of application is displaced a distance $\mathbf{D}$. The work is $(\mathbf{F}\cdot\mathbf{D}) = FD\cos\theta = F_xD_x + F_yD_y + F_zD_z$.

An *angular velocity* about an axis may be regarded as a vector. The direction of this vector is taken to be that in which an ordinary right-handed screw would move along the axis when rotating with the angular velocity.

The *outer product*, or *vector product*, of two vectors $\mathbf{V}$ and $\mathbf{U}$ is defined to be a vector of magnitude $VU\sin\theta$, where θ is the angle between $\mathbf{V}$ and $\mathbf{U}$, and direction that of the rotation from $\mathbf{V}$ to $\mathbf{U}$, through the angle θ. It is denoted by $[\mathbf{V}\cdot\mathbf{U}]$. Thus $[\mathbf{V}\cdot\mathbf{U}]$ is a vector equal to the area of the parallelogram formed by $\mathbf{V}$ and $\mathbf{U}$, and its direction is perpendicular to the plane containing $\mathbf{V}$ and $\mathbf{U}$. The angle θ is taken to be the positive angle less than 180° between $\mathbf{V}$ and $\mathbf{U}$. The vector $[\mathbf{V}\cdot\mathbf{U}]$ is of opposite sign to the vector $[\mathbf{U}\cdot\mathbf{V}]$, since the direction of rotation is reversed, so that

$$[\mathbf{V}\cdot\mathbf{U}] + [\mathbf{U}\cdot\mathbf{V}] = 0.$$

As an example of a vector product we may take the force per unit length on a current $\mathbf{C}$ in a magnetic field $\mathbf{H}$, which is equal to $CH\sin\theta$, and is perpendicular to $\mathbf{C}$ and $\mathbf{H}$. It is equal to $[\mathbf{C}\cdot\mathbf{H}]$.

The components of the vector product are equal to the projections of the area of the parallelogram formed by $\mathbf{V}$ and $\mathbf{U}$ on the planes perpendicular to the axes. This projection on the xy plane is equal to the area of the parallelogram formed by the points $(0, 0)$, (U_x, U_y), and (V_x, V_y), which is easily seen to be equal to $V_xU_y - V_yU_x$. The components of the vector product are therefore given by

$$\left.\begin{aligned}[\mathbf{V}\cdot\mathbf{U}]_x &= V_yU_z - V_zU_y,\\ [\mathbf{V}\cdot\mathbf{U}]_y &= V_zU_x - V_xU_z,\\ [\mathbf{V}\cdot\mathbf{U}]_z &= V_xU_y - V_yU_x.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Another example of a vector product is the moment of a force $\mathbf{F}$ acting at a point P with co-ordinates x, y, z about an axis through the origin

perpendicular to the plane containing the force and the origin. This is equal to $Fr\sin\theta$, where $r = OP$, and θ is the angle between OP produced and **F**. The x component of this vector product is the moment of **F** about the x axis, which is equal to $yF_z - zF_y$, so that

$$yF_z - zF_y = [\mathbf{r}\cdot\mathbf{F}]_x.$$

If **U**, **V**, **W** are three vectors then $(\mathbf{U}\cdot[\mathbf{V}\cdot\mathbf{W}])$ can easily be seen to be the volume of the parallelepiped formed by the three vectors. Hence if $(\mathbf{U}\cdot[\mathbf{V}\cdot\mathbf{W}]) = 0$ the three vectors must lie in the same plane

3. Scalar and Vector Fields.

The *field* of a vector or scalar quantity is a region throughout which at every point the quantity has a definite value, which in general varies from point to point. For example, the space around electrically charged bodies is a *vector field* of the electric field strength, and the atmosphere is a *scalar field* of the density of the air.

In the field of a scalar S the vector called the *gradient* of S is defined to be a vector the components of which are $\frac{\partial S}{\partial x}$, $\frac{\partial S}{\partial y}$, and $\frac{\partial S}{\partial z}$; or, if at any point in the field we draw a surface on which S is constant, then the gradient of S is normal to the surface, and its magnitude is the rate of variation of S per unit length along the normal. If p denotes the distance measured along a curve which is normal to the surfaces of constant S, then grad $S = \frac{\partial S}{\partial p}$, and $(\text{grad } S)_x = \frac{\partial S}{\partial x}$, $(\text{grad } S)_y = \frac{\partial S}{\partial y}$, $(\text{grad } S)_z = \frac{\partial S}{\partial z}$, so that we have the vector equation, grad $S = \frac{\partial S}{\partial p}$, and also

$$|\text{grad } S|^2 = \left(\frac{\partial S}{\partial p}\right)^2 = \left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 + \left(\frac{\partial S}{\partial z}\right)^2.$$

It is evident that the component of grad S in any direction is equal to $\frac{\partial S}{\partial s}$, where ∂s is an infinitesimal displacement in the given direction.

4. Line Integrals. Potentials.

Consider a curve of any shape drawn in a scalar or vector field. Let the distance measured along this curve be s, so that ds is an element of the curve. If Q denotes any quantity which has a definite value at every point in the field, then the sum of the products of the lengths of the elements ds of the curve and the values of Q at these elements, or $\int Q\,ds$, is called the line integral of Q along the curve. The line integral

for the part of the curve between two points A and B on it is denoted by $\int_A^B Q\,ds$. If $\mathbf{V}$ is a vector, and V_s denotes its component along the curve, then $\int_A^B V_s\,ds$ is the line integral of V_s along the curve. The product $V_s\,ds$ is equal to the scalar product of $\mathbf{V}$ and $\mathbf{ds}$, so that

$$\int_A^B V_s\,ds = \int_A^B (\mathbf{V} \cdot \mathbf{ds}) = \int_A^B (V_x\,dx + V_y\,dy + V_z\,dz),$$

where dx, dy, dz are the components of $\mathbf{ds}$. For the gradient of a scalar S, or grad S, we have therefore

$$\int_A^B (\text{grad } S)_s\,ds = \int_A^B \frac{\partial S}{\partial s}\,ds,$$

so that

$$\int_A^B (\text{grad } S)_s\,ds = S_B - S_A.$$

Thus the value of the integral $\int_A^B (\text{grad } S_s)\,ds$ between two points A and B depends only on the values of S at A and at B, and is independent of the position of the curve joining A and B along which the integral is taken.

If a vector $\mathbf{V}$ is equal to ($-$ grad S), where S is a scalar, then S is called the potential of V. Hence

$$\int_A^B V_s\,ds = -\int_A^B (\text{grad } S)_s\,ds = S_A - S_B,$$

that is, the potential difference between two points in the vector field is equal to the line integral of the vector component along any curve joining the points. The components of the vector are equal to the components of ($-$ grad S). so that

$$V_x = -\frac{\partial S}{\partial x}, \quad V_y = -\frac{\partial S}{\partial y}, \quad V_z = -\frac{\partial S}{\partial z}.$$

Thus if we know the potential of a vector at all points in the field we can easily calculate the vector at any point.

Examples of potentials are electrostatic potential, magnetostatic potential, and gravitational potential.

5. Vector Lines and Tubes.

A line drawn in the field of a vector $\mathbf{V}$ so as to be everywhere along the direction of $\mathbf{V}$ is called a *vector line*.

If we take a very small closed curve in the field and draw vector lines through all the points on it these lines enclose a tube called a *vector tube*.

The magnitude of the vector at a particular cross-section of a vector tube can be represented by drawing vector lines inside the tube

uniformly distributed over the cross-section, so that the number of the lines is proportional to the product of the magnitude of the vector and the area of the cross-section. Thus if a denotes the area of a cross-section of the tube and V the magnitude of the vector at this cross-section, then we suppose the number of vector lines drawn in the tube to be proportional to aV. The number of lines per unit area of cross-section is then proportional to V. In a very important class of cases the number of lines is constant from one cross-section to another, but this is not so in general (see below).

Instead of supposing the magnitude of the vector represented in this way by the number of vector lines drawn in the field, we may suppose the vector tube divided into a number of equal smaller tubes proportional to the product aV, and so have the number of tubes per unit area proportional to the magnitude of the vector. In this case V is inversely proportional to the cross-section of the vector tubes.

If the tubes are drawn so that the number per unit area on a surface perpendicular to the vector $\mathbf{V}$ is equal to the magnitude of $\mathbf{V}$, the tubes are called unit tubes of the vector $\mathbf{V}$. The cross-section of a unit tube at any point is then equal to $1/V$. The field is not well represented in this way unless the tubes are of small cross-section, but, of course, we may if necessary divide the unit tubes into as many smaller tubes as we like each of cross-section $1/nV$, where n is any convenient number.

The unit tubes may be closed tubes, or they may start at one point and end at another in the field. The number of unit tubes which start per unit volume at a point in the field minus the number which end per unit volume at the point is called the divergence of the vector $\mathbf{V}$ and is written div $\mathbf{V}$. The divergence of $\mathbf{V}$ may also be defined by the equation

$$\operatorname{div} \mathbf{V} = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z}. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

That the two definitions are equivalent will be shown in the next section.

6. Green's Theorem.

An important theorem giving transformations of volume integrals into surface integrals was discovered by Green. A special case discovered earlier by Gauss is known as Gauss's theorem.

Consider a closed surface of area σ enclosing a volume S. Let A be any function of position which has a definite value at all points inside the closed surface, and consider the integral $\int \frac{\partial A}{\partial x}\, dx\,dy\,dz$ taken

over the volume S. Draw a line through the volume S parallel to the x axis and let it cut the closed surface at two points P and Q. The line integral $\int \frac{\partial A}{\partial x}\,dx$ along this line from P to Q is equal to $A_Q - A_P$. If we multiply this by the small rectangular area $dy\,dz$ we get therefore the volume integral of $\partial A/\partial x$ over the volume of a rod of uniform cross-section $dy\,dz$ and length $x_Q - x_P$. Let $d\sigma$ be the area cut off on the surface σ by the rod at Q. Then if l denotes the cosine of the angle between the outward drawn normal to σ at Q and the x axis we have $dy\,dz = l\,d\sigma$, since $dy\,dz$ is equal to the projection of $d\sigma$ on the yz plane. Hence $A_Q\,dy\,dz = lA_Q\,d\sigma$. In the same way if $d\sigma'$ is the area cut off at P and l' the cosine of the angle between the outward normal to σ at P and the x axis, then

$$A_P\,dy\,dz = -l'A_P\,d\sigma'.$$

Hence

$$(A_Q - A_P)\,dy\,dz = lA_Q\,d\sigma + l'A_P\,d\sigma'.$$

Now the whole volume S may be divided up into small elements like $(x_Q - x_P)\,dy\,dz$, and for each of these the volume integral of $\partial A/\partial x$ is equal to an expression like that just found. We see therefore that the volume integral of $\partial A/\partial x$ for the whole volume S is equal to the sum of all these quantities like $lA_Q\,d\sigma + l'A_P\,d\sigma'$. But this sum is equal to the integral of $lA\,d\sigma$ over the whole surface σ, since when the whole volume S is divided into rod elements the ends of the elements will cover the whole surface σ. Hence we get

$$\int_S \frac{\partial A}{\partial x}\,dx\,dy\,dz = \int_\sigma lA\,d\sigma.$$

In the same way, if m and n are the cosines of the angle between an outward drawn normal to σ and the y and z axes, and B and C any functions of position inside the volume S, then

$$\int_S \frac{\partial B}{\partial y}\,dx\,dy\,dz = \int_\sigma mB\,d\sigma,$$

and

$$\int_S \frac{\partial C}{\partial z}\,dx\,dy\,dz = \int_\sigma nC\,d\sigma.$$

Adding the three results we get

$$\int_S \left(\frac{\partial A}{\partial x} + \frac{\partial B}{\partial y} + \frac{\partial C}{\partial z}\right)dx\,dy\,dz = \int_\sigma (lA + mB + nC)\,d\sigma, \quad . \quad (1)$$

which is Green's theorem.

If
$$A = V_x, \quad B = V_y, \quad C = V_z,$$

where $\mathbf{V}$ is a vector, then

$$lV_x + mV_y + nV_z = V_n,$$

where V_n denotes the component of $\mathbf{V}$ along the outward normal to $d\sigma$, so that

$$\int_S\left(\frac{\partial V_x}{\partial x}+\frac{\partial V_y}{\partial y}+\frac{\partial V_z}{\partial z}\right)dx\,dy\,dz=\int_\sigma V_n\,d\sigma. \quad . \quad . \quad (5)$$

Now since the normal cross-section of a unit vector tube is $1/\bar{V}$, the section by a surface the normal to which makes an angle θ with V is $1/V\cos\theta$. Thus, if $V_n = V\cos\theta$, we see that the number of unit tubes leaving the closed surface through the element of area $d\sigma$ is $V\cos\theta\, d\sigma = V_n\, d\sigma$. Hence $\int V_n d\sigma$ is equal to the total number of unit vector tubes coming out of the volume S through the surface σ. Tubes entering the surface are here understood to be reckoned negative.

Thus it appears that

$$\int_S\left(\frac{\partial V_x}{\partial x}+\frac{\partial V_y}{\partial y}+\frac{\partial V_z}{\partial z}\right)dx\,dy\,dz$$

is equal to the total number of unit vector tubes leaving the volume S. This is true for any volume, so that if we consider a very small one we see that

$$\frac{\partial V_x}{\partial x}+\frac{\partial V_y}{\partial y}+\frac{\partial V_z}{\partial z}$$

is equal to the number of unit tubes which start from it per unit volume. We have therefore

$$\operatorname{div}\mathbf{V}=\frac{\partial V_x}{\partial x}+\frac{\partial V_y}{\partial y}+\frac{\partial V_z}{\partial z}, \quad . \quad . \quad . \quad . \quad (3')$$

and so, putting $dx\,dy\,dz = dS$, we get

$$\int_S \operatorname{div}\mathbf{V}\,dS=\int_\sigma V_n\,d\sigma. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

This result is often called Gauss's theorem.

Green's theorem enables a volume integral of any function of x, y, and z which can be expressed in the form

$$\frac{\partial A}{\partial x}+\frac{\partial B}{\partial y}+\frac{\partial C}{\partial z}$$

to be transformed into a surface integral over the surface enclosing the volume.

The volume S may be the space between two closed surfaces, one inside the other, in which case the surface integral is taken over both

surfaces, and the normals are drawn in both cases in the direction *from* the enclosed volume *to* the surface.

7. Another Form of Green's Theorem.

Let
$$A=\phi\frac{\partial\psi}{\partial x},\quad B=\phi\frac{\partial\psi}{\partial y},\quad C=\phi\frac{\partial\psi}{\partial z},$$
so that
$$\frac{\partial A}{\partial x}+\frac{\partial B}{\partial y}+\frac{\partial C}{\partial z}=\phi\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right)+\frac{\partial\phi}{\partial x}\frac{\partial\psi}{\partial x}+\frac{\partial\phi}{\partial y}\frac{\partial\psi}{\partial y}+\frac{\partial\phi}{\partial z}\frac{\partial\psi}{\partial z}.$$

If $\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}$ be denoted by $\Delta\psi$, Green's theorem now gives
$$\int_S(\phi\Delta\psi+(\text{grad}\,\phi\cdot\text{grad}\,\psi))\,dS$$
$$=\int_\sigma\phi\left(l\frac{\partial\psi}{\partial x}+m\frac{\partial\psi}{\partial y}+n\frac{\partial\psi}{\partial z}\right)d\sigma=\int_\sigma\phi\frac{\partial\psi}{\partial n}\,d\sigma.$$

In the same way, interchanging ϕ and ψ, we get
$$\int_S(\psi\Delta\phi+(\text{grad}\,\phi\cdot\text{grad}\,\psi))\,dS=\int_\sigma\psi\frac{\partial\phi}{\partial n}\,d\sigma.$$

Thus, by subtraction,
$$\int_S(\phi\Delta\psi-\psi\Delta\phi)\,dS=\int_\sigma\left(\phi\frac{\partial\psi}{\partial n}-\psi\frac{\partial\phi}{\partial n}\right)d\sigma. \quad . \quad . \quad (7)$$

8. Solution of Poisson's Equation.

As an example of the use of this form of Green's theorem, we may find the solution of the equation
$$\frac{\partial^2 A}{\partial x^2}+\frac{\partial^2 A}{\partial y^2}+\frac{\partial^2 A}{\partial z^2}=\omega, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

or $\Delta A=\omega$, where A and ω are scalar functions of x, y, and z.

Put $\psi=A$ and $\phi=1/r$, where $r^2=x^2+y^2+z^2$, so that
$$\Delta\phi=-\frac{3}{r^3}+\frac{3(x^2+y^2+z^2)}{r^5}=0.$$

Hence
$$\int_S\frac{\omega}{r}\,dS=\int_\sigma\left(\frac{1}{r}\frac{\partial A}{\partial n}+\frac{A}{r^2}\frac{\partial r}{\partial n}\right)d\sigma.$$

Now let S be the volume between two concentric spheres of radii r_1 and r_2, with centre at the origin. In this case $\frac{\partial A}{\partial n}=\frac{\partial A}{\partial r}$ on the outer

sphere r_2, and $\frac{\partial A}{\partial n} = -\frac{\partial A}{\partial r}$ on the inner one r_1. Also $\frac{\partial r}{\partial n} = 1$ on the outer, and -1 on the inner, sphere. Therefore

$$\int_S \frac{\omega}{r} dS = \int_{\sigma_2} \left(\frac{1}{r_2} \frac{\partial A}{\partial r} + \frac{A}{r_2^2}\right) d\sigma - \int_{\sigma_1} \left(\frac{1}{r_1} \frac{\partial A}{\partial r} + \frac{A}{r_1^2}\right) d\sigma.$$

Now let the outer sphere be very large, and assume that A and $\partial A/\partial r$ become very small when r becomes very great, in such a way that the surface integral over the outer sphere tends to zero when r_2 is indefinitely increased. Also let the inner sphere be very small, so that A may be considered to be constant over its surface. Then

$$\int_S \frac{\omega}{r} dS = -4\pi r_1^2 \left(\frac{1}{r_1} \frac{\overline{\partial A}}{\partial r} + \frac{A}{r_1^2}\right),$$

where $\frac{\overline{\partial A}}{\partial r}$ denotes the average value of $\partial A/\partial r$ over the small sphere. Taking the limit for $r_1 = 0$, we have

$$\int_S \frac{\omega}{r} dS = -4\pi A,$$

or

$$A = -\frac{1}{4\pi}\int_S \frac{\omega}{r} dS. \quad \ldots \ldots \quad (9)$$

Here $-\frac{1}{4\pi}\int \frac{\omega}{r} dS$ is the value of A at the origin $r = 0$. The origin, of course, can be taken to be anywhere, so that we see that if

$$\Delta A = \omega,$$

then the value of A at any point can be obtained by dividing the space around the point into small elements of volume dS and multiplying each element by the value of ω/r at the element, where r is the distance of the element dS from the point at which A is required. The sum of all these products multiplied by $-1/4\pi$, i.e. $-\frac{1}{4\pi}\int \frac{\omega dS}{r}$, is then equal to A.

If, for example, $\omega = 0$ everywhere except in a small volume v in which its average value is $\bar{\omega}$, and if $v\bar{\omega} = e$, then at any point at a distance r from this small volume we have $A = -e/4\pi r$.

In the equations $\Delta A = \omega$ and $A = -\frac{1}{4\pi}\int \frac{\omega}{r} dS$, A and ω may be the components of vector quantities. Thus if $\mathbf{A}$ and $\boldsymbol{\omega}$ are vectors, and $\Delta \mathbf{A} = \boldsymbol{\omega}$, then

$$A_x = -\frac{1}{4\pi}\int \frac{\omega_x}{r} dS,$$

with similar equations for A_y and A_z.

Thus we see that $\Delta \mathbf{A} = \boldsymbol{\omega}$, where $\mathbf{A}$ and $\boldsymbol{\omega}$ are vectors, gives

$$\mathbf{A} = -\frac{1}{4\pi}\int \frac{\boldsymbol{\omega}\, dS}{r}$$

where $\int \frac{\boldsymbol{\omega}\, dS}{r}$ now denotes the vector sum of the products $\boldsymbol{\omega}\, dS/r$. We may regard the vector field of $\mathbf{A}$ as excited or produced by the quantity $\boldsymbol{\omega}$.

9. Solution of Equation for a Propagated Potential.

Consider next the equation

$$\Delta A - \frac{1}{c^2}\ddot{A} = \omega, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where c is a constant and $\ddot{A} = \partial^2 A/\partial t^2$. This important equation can be regarded as meaning that ω excites a field which is propagated out from ω with the velocity c.

The solution of this equation can be obtained by means of Green's theorem in a manner similar to that used to get the solution of $\Delta A = \omega$.

Put $\phi = \frac{1}{r} F(t + r/c)$, and $\psi = A$, in equation (7):

$$\int_S (\phi\Delta\psi - \psi\Delta\phi)\, dS = \int_\sigma \left(\phi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\phi}{\partial n}\right) d\sigma.$$

Here $F(t + r/c)$ denotes an arbitrary function of $t + r/c$, which can be given any desired form. It can easily be shown that $\phi = \frac{1}{r}F(t + r/c)$ satisfies $\Delta\phi - \frac{1}{c^2}\ddot{\phi} = 0$, by substituting the values of the differential coefficients of ϕ in this equation. We have therefore

$$\int_S \left(\phi\left(\omega + \frac{1}{c^2}\ddot{A}\right) - A\frac{1}{c^2}\ddot{\phi}\right) dS = \int_\sigma \left(\phi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\phi}{\partial n}\right) d\sigma,$$

or
$$\int_S \phi\omega\, dS + \frac{d}{dt}\int_S \frac{(\phi\dot{A} - A\dot{\phi})\, dS}{c^2} = \int_\sigma \left(\phi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\phi}{\partial n}\right) d\sigma.$$

Multiply by dt and integrate from t_1 to t_2. Thus

$$\int_{t_1}^{t_2} dt \int_S \phi\omega\, dS + \left[\int_S \frac{(\phi\dot{A} - A\dot{\phi})\, dS}{c^2}\right]_{t_1}^{t_2} = \int_{t_1}^{t_2} dt \int_\sigma \left(\phi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\phi}{\partial n}\right) d\sigma. \quad (11)$$

Now suppose that $F(t + r/c) = 0$ except for values of $t + r/c$ very nearly $= 0$. Also let t_2 be greater than $-r/c$, and t_1 less, so that

the instant $t = -r/c$ is between t_1 and t_2. Also suppose that $\int_{t_1}^{t_2} F(t+r/c)\,dt = 1$. This requires that $F(0)$ be so large that although $F(t+r/c)$ is zero for all values of $t+r/c$, except those near zero, yet $\int_{t_1}^{t_2} F(t+r/c)\,dt = 1$. The function $F(t+r/c)$ is thus made to act as a selecting factor, since when multiplied by any quantity and integrated over the time it selects the value of the quantity at the time $t = -r/c$.

Thus

$$\int_{t_1}^{t_2} dt \int_S \phi\omega\,dS = \int \frac{[\omega]}{r}\,dS,$$

where $[\omega]$ denotes the value of ω in the element of volume dS at the time $t = -r/c$.

Hence
$$\int_S \frac{[\omega]}{r}\,dS = \int_{t_1}^{t_2} dt \int_\sigma \left(\phi\frac{\partial A}{\partial n} - A\frac{\partial\phi}{\partial n}\right) d\sigma,$$

the second term on the left of equation (11) disappearing, since ϕ and $\dot{\phi}$ vanish when $t = t_1$ and $t = t_2$. Now let the boundary of S be a very small sphere of radius r_1 and a large sphere of radius r_2, both with centre at the origin. Assume that A and $\partial A/\partial n$ are zero on the large sphere, so that only the integral over the surface of the small sphere need be considered.

We have $\dfrac{\partial A}{\partial n} = -\dfrac{\partial A}{\partial r}$ and $\dfrac{\partial\phi}{\partial n} = -\dfrac{\partial\phi}{\partial r}$; also

$$\frac{\partial\phi}{\partial r} = \frac{\partial}{\partial r}\left(\frac{1}{r}F(t+r/c)\right) = -\frac{1}{r_1^2}F + \frac{1}{r_1 c}F',$$

where F' is the differential coefficient of F with respect to $t+r/c$. Hence
$$\phi\frac{\partial A}{\partial n} - A\frac{\partial\phi}{\partial n} = -\frac{F}{r_1}\frac{\partial A}{\partial r} - \frac{AF}{r_1^2} + \frac{AF'}{r_1 c}.$$

When r_1 is very small, we may regard A and F as constant over the surface of the small sphere, and replace $\partial A/\partial r$ and F' by their average values, so that

$$\int_{t_1}^{t_2} dt \int_\sigma \left(\phi\frac{\partial A}{\partial n} - A\frac{\partial\phi}{\partial n}\right) d\sigma = \int_{t_1}^{t_2} 4\pi r_1^2\left(\frac{\overline{AF'}}{r_1 c} - \frac{AF}{r_1^2} - \frac{F}{r_1}\frac{\partial A}{\partial r}\right) dt.$$

In the limit when $r_1 = 0$, this becomes

$$-4\pi\int_{t_1}^{t_2} AF\,dt,$$

which is equal to the value of $-4\pi A$ at the time $t = 0$, because F is zero except at $t = -r/c$, i.e. when $r = 0$, at $t = 0$.

Hence we have finally

$$A_{t=0} = -\frac{1}{4\pi}\int_S \frac{[\omega]\,dS}{r}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

where $[\omega]$ denotes the value of ω in the element of volume dS at the time $t = -r/c$, and r is the distance from dS to the point at which A is to be calculated.

Thus we see that the field at any time due to the ω in any element of volume is determined by the value of ω at a time r/c previously, so that the field excited by ω travels out with the velocity c.

10. Curl of a Vector.

In the field of any vector $\mathbf{V}$ we can define another vector $\mathbf{R}$ determined by the variation of $\mathbf{V}$ in the following way. Consider a small plane area α enclosed by a curve of length s. Take the line integral $\int V_s\,ds$ of the component V_s of $\mathbf{V}$ along s round the curve s. Then $R_n = \dfrac{1}{\alpha}\int_s \mathbf{V}_s\,ds$ can be proved to be the component of a vector $\mathbf{R}$ along the normal n to the plane area α. The direction of $\mathbf{R}$ along the normal to α is taken so that a right-handed screw would advance in the direction of $\mathbf{R}$ when turned in the direction in which the line integral round s is taken. The vector $\mathbf{R}$ is called the *curl* (or *rotation*) of $\mathbf{V}$ and is denoted by curl $\mathbf{V}$ (or rot $\mathbf{V}$).

11. Stokes's Theorem.

We shall now consider an important theorem due to Sir G. G. Stokes, which enables an integral over a surface to be transformed into a line integral round the boundary of the surface.

Consider a surface of any shape in the field of a vector $\mathbf{V}$ and suppose a closed curve drawn on it. Let s denote the length of the closed curve and σ the area on the surface enclosed by the curve.

Let $d\sigma$ be an element of area on the surface, n the normal to the surface at $d\sigma$, and ds an element of the closed curve. Then Stokes's theorem is:

$$\int_\sigma (\text{curl}\,\mathbf{V})_n\,d\sigma = \int_s V_s\,ds, \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

or, in words, the integral of the normal component of the curl of a vector over the surface, is equal to the line integral round the boundary of the component of the vector along the bounding curve.

Let ABC be the closed curve, and let the surface bounded by it be divided up into a large number of small elements by two sets of curves, as shown in fig. 1.

Consider the sum of the line integrals of V_s round the boundaries

of all these elements going round them all the same way. Each line element not on the outside boundary will appear twice in this sum, once for each of the two elements of area between which it lies. The two line integrals along any line element not on the boundary are equal but of opposite sign, since the element is traversed in opposite directions when going round the two elements of area on either side of the line. The sum of the line integrals of V_s round all the elements of area into which the whole area is divided is therefore equal to the line integral of V_s round the outside boundary, or $\int_s V_s\,ds$.

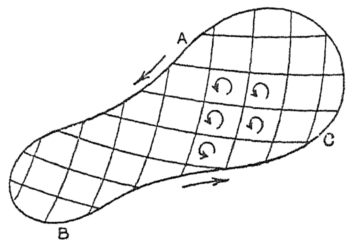


Fig 1

Now, for any element of area $d\sigma$, the line integral of V_s round its boundary is equal to the component of curl **V** along the normal to the element multiplied by $d\sigma$, or $(\text{curl}\,\mathbf{V})_n\,d\sigma = \int V_s\,ds$. Hence the sum of all the line integrals round the elements of area is equal to

$$\int_\sigma (\text{curl}\,\mathbf{V})_n\,d\sigma,$$

so that we get

$$\int_\sigma (\text{curl}\,\mathbf{V})_n\,d\sigma = \int_s V_s\,ds,$$

which is Stokes's theorem.

12. **Components of the Curl of a Vector.**

The components of the vector curl **V** relative to axes x, y, z can be easily calculated. Consider a small rectangular area parallel to the yz plane with sides dy and dz. The normal to it is along the x direction, so that we have

$$(\text{curl}\,\mathbf{V})_x = \frac{1}{dy\,dz}\int V_s\,ds,$$

the line integral of V_s being taken round the rectangular area. Hence

$$\int V_s\,ds = (V_y - V_y')\,dy + (V_z' - V_z)\,dz,$$

where V_y, V_z and V_y', V_z' are the values on opposite sides of the area $dy\,dz$.

But $V_y' - V_y = \dfrac{\partial V_y}{\partial z}dz$, and $V_z' - V_z = \dfrac{\partial V_z}{\partial y}dy$, so that

$$\int V_s\,ds = \left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z}\right)dy\,dz.$$

and therefore

$$\left.\begin{aligned}(\text{curl } \mathbf{V})_x &= \frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z}, \\ (\text{curl } \mathbf{V})_y &= \frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x}, \\ (\text{curl } \mathbf{V})_z &= \frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y}.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

The curl of a vector may be defined by these equations instead of (§10) by

$$(\text{curl } \mathbf{V})_n = \frac{1}{a}\int V_s\, ds. \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

13. **Some Important Relations.**

We have

$$\text{div}\,(\text{curl } \mathbf{V}) = \frac{\partial}{\partial x}\left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y}\right)$$

$$= \frac{\partial^2 V_z}{\partial y\,\partial z} - \frac{\partial^2 V_y}{\partial z\,\partial y} + \frac{\partial^2 V_x}{\partial z\,\partial x} - \frac{\partial^2 V_z}{\partial x\,\partial z} + \frac{\partial^2 V_y}{\partial x\,\partial y} - \frac{\partial^2 V_x}{\partial y\,\partial x},$$

or

$$\text{div}\,(\text{curl } \mathbf{V}) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

This means that the vector lines of curl **V** always form closed curves, so that the number of unit tubes of curl **V** crossing any section of a vector tube of curl **V** is constant along the tube. Hence if a is the area of the cross-section of a vector tube of curl **V**, a curl **V** is constant along the tube. This is true for any vector the divergence of which is zero; it follows at once from Green's theorem (5).

If the curl of a vector **V** is zero throughout a region in the field of the vector, the line integral of V_s round any closed curve s in this region is zero. The vector therefore has a potential, or is equal to the negative gradient of a scalar. The converse is proved as follows.

$$(\text{curl}\,(\text{grad } S))_x = \frac{\partial}{\partial y}(\text{grad } S)_z - \frac{\partial}{\partial z}(\text{grad } S)_y$$

$$= \frac{\partial^2 S}{\partial y\,\partial z} - \frac{\partial^2 S}{\partial z\,\partial y} = 0,$$

with similar equations for the y and z components, so that

$$\text{curl}\,(\text{grad } S) = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

The field of a vector the curl of which is zero is said to be *irrotational*, and that of a vector with zero divergence is called *solenoidal*.

The curl of the curl of a vector is an important quantity. We have

$$
\begin{aligned}
(\text{curl}\,(\text{curl}\,\mathbf{V}))_x &= \frac{\partial}{\partial y}(\text{curl}\,\mathbf{V})_z - \frac{\partial}{\partial z}(\text{curl}\,\mathbf{V})_y \\
&= \frac{\partial^2 V_y}{\partial y \partial x} - \frac{\partial^2 V_x}{\partial y^2} - \frac{\partial^2 V_x}{\partial z^2} + \frac{\partial^2 V_z}{\partial z\, \partial x} \\
&= \frac{\partial}{\partial x}\left(\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z}\right) - \left(\frac{\partial^2 V_x}{\partial x^2} + \frac{\partial^2 V_x}{\partial y^2} + \frac{\partial^2 V_x}{\partial z^2}\right) \\
&= (\text{grad div}\,\mathbf{V})_x - \Delta V_x,
\end{aligned}
$$

with similar equations for the y and z components. Hence

$$\text{curl curl}\,\mathbf{V} = \text{grad div}\,\mathbf{V} - \Delta\mathbf{V}. \quad . \quad . \quad . \quad . \quad (18)$$

We have seen (§8) that if $\Delta\mathbf{V} = \boldsymbol{\omega}$, then

$$\mathbf{V} = -\frac{1}{4\pi}\int\frac{\boldsymbol{\omega}\,dS}{r},$$

so that, since

$$\Delta\mathbf{V} = \text{grad div}\,\mathbf{V} - \text{curl curl}\,\mathbf{V}$$

we have, for any vector $\mathbf{V}$,

$$\mathbf{V} = \frac{1}{4\pi}\int\frac{1}{r}\,\text{curl curl}\,\mathbf{V}\,dS - \frac{1}{4\pi}\int\frac{1}{r}\,\text{grad div}\,\mathbf{V}\,dS. \quad . \quad . \quad (19)$$

If the divergence of a vector $\mathbf{V}$ is equal to zero, it is always possible to find another vector $\mathbf{A}$ such that $\mathbf{V} = \text{curl}\,\mathbf{A}$. For if $\mathbf{A} = \frac{1}{4\pi}\int\frac{1}{r}\,\text{curl}\,\mathbf{V}\,dS$, then (19) gives $\mathbf{V} = \text{curl}\,\mathbf{A}$.

In the same way, if curl $\mathbf{V} = 0$, it is always possible to find a scalar ϕ such that $\mathbf{V} = -\text{grad}\,\phi$. Cf. (17) above.

The divergence of the vector product of two vectors can be easily calculated. We have

$$
\begin{aligned}
\text{div}\,[\mathbf{A}\cdot\mathbf{B}] &= \frac{\partial}{\partial x}[A\cdot B]_x + \frac{\partial}{\partial y}[A\cdot B]_y + \frac{\partial}{\partial z}[A\cdot B]_z \\
&= \frac{\partial}{\partial x}\left(A_y B_z - A_z B_y\right) + \frac{\partial}{\partial y}\left(A_z B_x - A_x B_z\right) + \frac{\partial}{\partial z}(A_x B_y - A_y B_x) \\
&= B_x\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) + B_y\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) + B_z\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) \\
&\quad - A_x\left(\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z}\right) - A_y\left(\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x}\right) - A_z\left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right),
\end{aligned}
$$

or

$$\text{div}\,[\mathbf{A}\cdot\mathbf{B}] = (\mathbf{B}\cdot\text{curl}\,\mathbf{A}) - (\mathbf{A}\cdot\text{curl}\,\mathbf{B}). \quad . \quad . \quad (20)$$

11. Theorem of Coriolis.

The theory of the motion of a point relative to moving axes affords a good example of the use of vector analysis.

Consider any vector **V** represented by the line OP in the diagram. Suppose that in time t, where t is small, P moves to R so that the vector **V** changes from OP to OR.

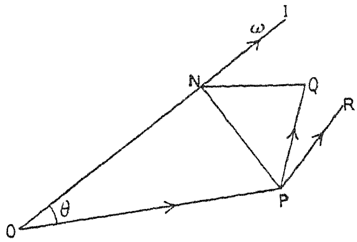


Fig. 2

Also suppose that we refer the motion of P to moving axes, and that if it remained fixed relatively to these moving axes it would move from P to Q in the time t. The displacement PQ therefore represents the motion of the axes during t. Let the motion of the axes be a rotation with angular velocity $\boldsymbol{\omega}$ about an axis OI through O. The position of the axis of rotation may change with the time, but we suppose that at the instant considered it is along OI. We have then $PQ = NP \cdot \omega t$, where NP is the perpendicular from P to OI. But $NP = V \sin\theta$, so that, since PQ is perpendicular to the plane containing OI and OP, we see that

$$PQ = [\boldsymbol{\omega} t \cdot \mathbf{V}].$$

The motion of P relative to the moving axes is represented by QR, and we have the vector equation

$$PR = PQ + QR,$$

or

$$PR = QR + [\boldsymbol{\omega} \cdot \mathbf{V}t].$$

Now PR represents the change in **V** in time t, so that we may put $PR/t = \dot{\mathbf{V}}$; and QR represents the change in **V** relative to the rotating axes, so that we may put $QR/t = \dot{\mathbf{V}}'$. Hence

$$\dot{\mathbf{V}} = \dot{\mathbf{V}}' + [\boldsymbol{\omega} . \mathbf{V}].$$

Here $\dot{\mathbf{V}}$ is the rate of change of any vector **V** relative to fixed axes, and $\dot{\mathbf{V}}'$ the rate of change relative to axes rotating with angular velocity $\boldsymbol{\omega}$.

Now consider the displacement **D** of a point P, so that

$$\dot{\mathbf{D}} = \dot{\mathbf{D}}' + [\boldsymbol{\omega} \cdot \mathbf{D}];$$

and also consider the velocity $\dot{\mathbf{D}}$ of the same point, so that

$$\ddot{\mathbf{D}} = \frac{d'}{dt}(\dot{\mathbf{D}}) + [\boldsymbol{\omega} \cdot \dot{\mathbf{D}}].$$

Here $\ddot{\mathbf{D}}$ is the acceleration of the point relative to fixed axes. Hence

$$\ddot{\mathbf{D}} = \frac{d'}{dt}(\dot{\mathbf{D}}' + [\boldsymbol{\omega} \cdot \mathbf{D}]) + [\boldsymbol{\omega} \cdot (\dot{\mathbf{D}}' + [\boldsymbol{\omega} \cdot \mathbf{D}])],$$

where d'/dt indicates the rate of variation relative to the moving axes. Hence

$$\ddot{\mathbf{D}} = \ddot{\mathbf{D}}' + [\dot{\boldsymbol{\omega}}' \cdot \mathbf{D}] + [\boldsymbol{\omega} \cdot \dot{\mathbf{D}}'] + [\boldsymbol{\omega} \cdot \dot{\mathbf{D}}'] + [\boldsymbol{\omega} \cdot [\boldsymbol{\omega} \cdot \mathbf{D}]],$$

so that $$\ddot{\mathbf{D}} = \ddot{\mathbf{D}}' + 2[\boldsymbol{\omega} \cdot \dot{\mathbf{D}}'] + [\dot{\boldsymbol{\omega}}' \cdot \mathbf{D}] + [\boldsymbol{\omega} \cdot [\boldsymbol{\omega} \cdot \mathbf{D}]]. \quad . \quad (21)$$

This result is sometimes called *Coriolis's theorem*, after its discoverer.

If the angular velocity $\boldsymbol{\omega}$ of the moving axes is constant in magnitude and direction,

$$\dot{\boldsymbol{\omega}}' = 0,$$

and $$\ddot{\mathbf{D}} = \ddot{\mathbf{D}}' + 2[\boldsymbol{\omega} \cdot \dot{\mathbf{D}}'] + [\boldsymbol{\omega} \cdot [\boldsymbol{\omega} \cdot \mathbf{D}]].$$

$[\boldsymbol{\omega} \cdot \mathbf{D}]$ is equal to $PN \cdot \omega$, and, since PN is perpendicular to $\boldsymbol{\omega}$, $[\boldsymbol{\omega} \cdot [\boldsymbol{\omega} \cdot \mathbf{D}]]$ is equal to $PN \cdot \omega^2$, and so is simply the centripetal acceleration due to the rotation. The term $2[\boldsymbol{\omega} \cdot \dot{\mathbf{D}}']$ is called the *Coriolis acceleration*. It represents an acceleration in a plane perpendicular to the axis of rotation. If the velocity $\dot{\mathbf{D}}'$ is directed away from the axis, in such a plane, then $2[\boldsymbol{\omega} \cdot \dot{\mathbf{D}}']$ is equal to $2\omega \dot{D}'$.

Equivalent results to those above are easily found by analytical methods. For example, if the axes of x and y are rotating about Oz with any angular velocity $\boldsymbol{\omega}$, the velocities of the point $P(x, y, z)$ parallel to Ox and Oy are

$$u = \dot{x} - y\omega, \quad v = \dot{y} + x\omega.$$

The accelerations of P are found from u and v in the same way as u and v are found from x and y. The component accelerations parallel to Ox and Oy are therefore

$$\dot{u} - v\omega = \ddot{x} - x\omega^2 - y\dot{\omega} - 2\dot{y}\omega,$$

and $$\dot{v} + u\omega = \ddot{y} - y\omega^2 + x\dot{\omega} + 2\dot{x}\omega.$$

The terms $-2\dot{y}\omega$, $2\dot{x}\omega$ are the components of the Coriolis acceleration.

15. Fourier's Series and Integrals.

A function of x, say $f(x)$, may be represented by the series

$$f(x) = \sum_{n=-\infty}^{n=+\infty} c_n e^{2\pi i n x/l},$$

where n is an integer, for values of x between 0 and l. To determine the c_n's multiply by $e^{-2\pi imx/l}dx$ and integrate from 0 to l: thus

$$\int_0^l f(x)e^{-2\pi imx/l}dx = \sum_{n=-\infty}^{n=+\infty} c_n \int_0^l e^{2\pi i(n-m)x/l}dx.$$

The integrals on the right are equal to zero when n is not equal to m but to l when $n = m$, so

$$c_m = \frac{1}{l}\int_0^l f(x)e^{-2\pi imx/l}dx.$$

If $f(x)$ is real, then $c_n e^{2\pi inx/l} + c_{-n}e^{-2\pi inx/l}$ must be real, so that $c_n = c_{-n}$ and

$$c_n e^{2\pi inx/l} + c_{-n}e^{-2\pi inx/l} = 2|c_n|\cos(2\pi nx/l + \theta).$$

The average value of the square of this is $2|c_n|^2$.

If $f(x)$ is zero for very large positive and negative values of x, then for values of x between $-\infty$ and $+\infty$ we may assume that

$$f(x) = \int_{-\infty}^{+\infty} c(n)e^{2\pi inx}dn,$$

where $c(n)$ is a function of n. To find $c(n)$ multiply by $e^{-2\pi imx}$ and integrate from $-\infty$ to $+\infty$, so that

$$\int_{-\infty}^{+\infty} f(x)e^{-2\pi imx}dx = \int\int c(n)e^{2\pi i(n-m)x}dx\,dn.$$

Now

$$\int_{-l}^{+l} e^{2\pi i(n-m)x}dx = \frac{\sin 2\pi(n-m)l}{\pi(n-m)},$$

so that we may take

$$\int_{-\infty}^{+\infty} f(x)e^{-2\pi imx}dx = \left[\int_{-\infty}^{+\infty} c(n)\frac{\sin 2\pi(n-m)l}{\pi(n-m)}dn\right]_{l=\infty}.$$

Let $z = 2\pi(n-m)$ and $dz = 2\pi dn$, which gives

$$\int_{-\infty}^{+\infty} f(x)e^{-2\pi imx}dx = \frac{1}{\pi}\left[\int_{-\infty}^{+\infty} c(n)\frac{\sin lz}{z}dz\right]_{l=\infty}.$$

The value of $\int_{z'}^{z'+\Delta}\frac{\sin lz}{z}dz$, when l tends to ∞ and z' and Δ are either both positive or both negative, is easily seen to tend to zero unless z' is very small, so that, since when $z = 0$ then $n = m$, we have

$$\int_{-\infty}^{+\infty} f(x)e^{-2\pi imx}dx = c(m)$$

because

$$\int_{-\infty}^{+\infty}\frac{\sin lz}{z}dz = \pi$$

for all positive values of l.

Putting $c(n) = \int_{-\infty}^{+\infty} f(v)e^{-2\pi i n v}dv$ in $f(x) = \int_{-\infty}^{+\infty} c(n)e^{2\pi i n x}dn$, we get

$$f(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{+\infty} f(v)e^{2\pi i n(x-v)}dv\,dn,$$

which is one form of Fourier's integral. These theorems are true for finite and continuous functions which have only a finite number of maxima and minima in any finite range. They hold also for functions with finite discontinuities except at the points of discontinuity.

16. Theory of Two Equal Coupled Oscillators.

Consider a system of two equal masses both oscillating along Ox and acted on by equal restoring forces $-\mu x$ and also by forces between the particles proportional to the distance between them. The equations of motion are then

$$m\ddot{x}_1 + \mu x_1 + a(x_1 - x_2) = 0,$$
$$m\ddot{x}_2 + \mu x_2 + a(x_2 - x_1) = 0.$$

Let $(\mu + a)/m = q^2$ and $a/m = c$, so that

$$\ddot{x}_1 + q^2x_1 - cx_2 = 0,$$
$$\ddot{x}_2 + q^2x_2 - cx_1 = 0.$$

If the system is oscillating with a definite frequency, we may suppose $x_1 = Ae^{ipt}$ and $x_2 = Be^{ipt}$, so that

$$-Ap^2 + q^2A - cB = 0,$$
$$-Bp^2 + q^2B - cA = 0.$$

Eliminating A/B, this gives $p^2 = q^2 \pm c$. If there were no force between the particles, we should have $p^2 = q^2$ and the frequency of the oscillation would be $q/2\pi$, so that the coupling changes the frequency from $q/2\pi$ to $\nu_1 = \sqrt{q^2 + c}/2\pi$ or $\nu_2 = \sqrt{q^2 - c}/2\pi$. The two equal coupled oscillators therefore have two possible frequencies, one greater and one less than the frequency without coupling.

If $p^2 = q^2 + c$, then $B = -A$, and if $p^2 = q^2 - c$, then $B = A$. Thus the two particles vibrate with the higher frequency with opposite phases and with the lower frequency with equal phases.

Any linear combination of the two solutions

$$x_1 = Ae^{2\pi i\nu_1 t},\ x_2 = -Ae^{2\pi i\nu_1 t}$$
$$x_1 = Ae^{2\pi i\nu_2 t},\ x_2 = Ae^{2\pi i\nu_2 t}$$

is also a solution, so we may put, for example,

$$x_1 = A(e^{2\pi i\nu_2 t} + e^{2\pi i\nu_1 t})$$
$$x_2 = A(e^{2\pi i\nu_2 t} - e^{2\pi i\nu_1 t})$$

or

$$x_1 = 2Ae^{\pi it(\nu_1+\nu_2)}\cos \pi t(\nu_1 - \nu_2)$$
$$x_2 = 2Ae^{\pi it(\nu_1+\nu_2)-i\pi/2}\sin \pi t(\nu_1 - \nu_2).$$

In this case, if c is small so that $(\nu_1 - \nu_2)/(\nu_1 + \nu_2)$ is small, we may say that x_1 oscillates with frequency $(\nu_1 + \nu_2)/2$ and slowly varying amplitude $2A\cos \pi t(\nu_1 - \nu_2)$, and x_2 oscillates with the same frequency and amplitude $2A\sin \pi t(\nu_1 - \nu_2)$. At $t = 0$, x_1 has amplitude $2A$ and x_2 has amplitude zero, and at $t = 1/2(\nu_1 - \nu_2)$, x_1 has amplitude zero and x_2 amplitude $2A$.

Thus if at $t = 0$, x_1 is oscillating and x_2 is not, then as time goes on the oscillations of x_1 die down and those of x_2 increase until x_2 is oscillating and x_1 not. The transfer of the oscillations from one particle to the other and back again goes on indefinitely.

REFERENCES

1. *Synopsis of Applicable Mathematics.* L. Silberstein.
2. *A Short Table of Integrals.* Peirce.
3. *Smithsonian Mathematical Formulæ.* E. P. Adams.
4. *The Mathematics of Physics and Chemistry.* Margenau and Murphy.

EXAMPLES

CHAPTER I

1. Show by means of the electromagnetic equations for empty space that in a plane electromagnetic wave the electric and magnetic fields are perpendicular to each other and to the direction of propagation of the wave.

2. By considering the electromagnetic momentum in a train of waves, show that the waves exert a pressure on a surface in which they are absorbed equal to the energy density in the waves.

3. A sphere of electricity of uniform density expands so that its radius increases while its total charge remains constant. Show that the current and the magnetic field in the sphere are everywhere equal to zero during the expansion.

4. An electron is projected with velocity v from a point at a distance d from a magnetic pole. If r denotes the distance of the electron from the pole after a time t, show that

$$r^2 = v^2t^2 + 2av_rt + d^2,$$

where v_r denotes the component of the velocity v along the direction of the magnetic field at $t = 0$.

5. Show that the electromagnetic angular momentum per unit length along the axis of a condenser consisting of two concentric cylinders of radii b and a is equal to $(b^2 - a^2)he/2c$, when the charge per centimetre on the inner cylinder is e and there is a magnetic field h along the axis.

6. Show that the force on a magnetic pole m moving with velocity v in an electric field F is equal to $Fm(v/c)\sin\theta$, where θ is the angle between F and v.

7. Show that the mean rate of radiation of energy from an electron describing a simple harmonic motion of amplitude a and frequency ν is equal to $\frac{4}{3}(\pi/c)^3 e^2a^2\nu^4$.

8. Show that

$$\frac{d}{dt}\left\{\frac{1}{2}\int_S(\mathbf{F}^2 + \mathbf{H}^2)\,dS\right\} + \int_S \rho(\mathbf{F}\cdot\mathbf{V})\,dS + c\int_\sigma [\mathbf{F}\,\mathbf{H}]_n\,d\sigma = 0,$$

where $\mathbf{F}$ and $\mathbf{H}$ are the electric and magnetic field strengths, ρ the density of electricity, $\mathbf{V}$ the velocity of the electricity, S is the volume enclosed by a surface of area σ, and the suffix n indicates the component along the normal to the element of area $d\sigma$.

9. A condenser consisting of two concentric cylinders with radii b and a, with the space between them filled with a medium of specific inductive capacity K and magnetic permeability μ, is rotating with angular velocity ω about its axis in a magnetic field H parallel to its axis. Show that if the cylinders are connected together there are equal and opposite charges on them equal to $\pm\,\pi\omega H(\mu K - 1)(b^2 - a^2)/c\log(b/a)$ per unit length along the axis.

10. A hollow circular cylinder of specific inductive capacity K is placed in a radial electric field of strength A/r at a distance r from the axis of the cylinder. If the cylinder is rotating with angular velocity ω, calculate the magnetic field in the cylinder.

CHAPTER II

1. Calculate the magnetic field at a point on the axis of a circle round which an electron is moving with uniform velocity. Show that its component along the axis is equal to the field due to a current, round the circle, equal to the charge on the electron multiplied by the number of revolutions it makes per second.

2. A magnetic field H is suddenly generated perpendicular to the plane of a uniform circular ring of electricity of charge e, mass m, and radius a. Calculate the resulting angular velocity of the ring, assuming the field due to the ring very small compared with H.

3. What would be the apparent magnetic permeability of a long uniform bar the material of which was a perfect conductor of electricity?

4. In the case of a large number of similar systems each having possible energies $0, \varepsilon, 2\varepsilon, 3\varepsilon, \ldots n\varepsilon \ldots$, show that in the equilibrium state the fraction for which the energy E is between E and $E + dE$ is equal to $\varepsilon^{-E/kT}\, dE/kT$ when ε is made indefinitely small.

5. If $\mathbf{I}$ is defined as the magnetic moment per unit volume, and $\mathbf{B} = \mathbf{H} + 4\pi\mathbf{I}$, where $\mathbf{H}$ is the magnetic field strength, defined as usual as the field inside a long narrow cavity parallel to $\mathbf{I}$, show that $\operatorname{div}\mathbf{B} = 0$.

6. A space in which there is a uniform magnetic field H contains n electrons per cubic centimetre all moving with velocities of the same magnitude v in planes perpendicular to H. Neglecting the mutual action of the electrons, calculate the intensity of magnetization. Ans. $\frac{1}{2}nmv^2/H$.

7. Discuss the analogy between the "critical temperature" at which ferromagnetic properties disappear and the "critical temperature" above which a vapour cannot be liquefied.

8. Calculate the magnetic field at the centre of the circular electron orbit of a Bohr magneton.

9. An electron of mass m, spinning so that it has a magnetic moment equal to that of a Bohr magneton, is projected with velocity v from a point at a distance d from a north pole P. If v_1 is its initial velocity along the direction of the field due to the pole, and if its magnetic axis points towards the pole, how far from the pole will it get? Ans. $\sqrt{[(9.2 \times 10^{-21}P)/(9.2 \times 10^{-21}P/d^2 - \frac{1}{2}mv_1^2)]}$.

10. In Stern and Gerlach's experiment, if $l = 5$ cm. and $dH/dy = 100{,}000$, calculate the mean deviation y for silver atoms having a temperature of 1000° C.

CHAPTER III

1. Show that, on the classical theory of thermionic emission, the mean kinetic energy of the electrons which escape from the metal is equal to that of the electrons inside the metal as they enter the surface layer.

2. The thermionic current i from a wire is found to be represented by the equation $i = AT^2\varepsilon^{-b/T}$, where A and b are constants. Show that it can be equally well represented by $i = A'T^2\varepsilon^{-b'/T}$, where A' and b' are not equal to A and b, provided that

$$T \log \frac{A}{A'} = b - b'.$$

3. If the latent heat of evaporation of the electrons from a metal is given by

$L = \dfrac{RT^2}{p}\dfrac{dp}{dT}$, where p is the pressure of the electron gas, and if $L = L_0 + aT$, where a is a constant, show that the thermionic current density is given by $i = AT^{a/R}\epsilon^{-L_0/RT}$.

4. Calculate the maximum thermionic current from a straight wire, 0·1 mm. in diameter and 10 cm. long, to a cylinder surrounding it, of 2 cm. radius, due to a potential difference of 100 volts.

5. What fraction of the saturation thermionic current from a wire at 1500° C. would be obtained with an electrode surrounding the wire at a potential of one volt above that of the wire?

CHAPTER IV

1. If a plane surface, at potential V, illuminated by ultra-violet light, emits electrons, equally in all directions and all with velocity v, from every point illuminated, show that the current to a parallel plane at potential zero is equal to $ne\{1 - \sqrt{(2Ve)/(mv^2)}\}$, where n is the total number of electrons emitted per second.

2. The potential difference between two large parallel metal plates is V. There is a uniform magnetic field H in the space between them parallel to the surfaces of the plates. A small area on the negative plate is illuminated and emits n electrons per second with negligible velocities. Calculate the current between the plates when n is so small that the space charge can be neglected.

3. If $\frac{1}{4}c\int_0^\infty T_\nu F(\nu)E(\nu)\,d\nu = 2NkT$ (p. 64), show that $T_\nu = h\nu - w_0$.

4. Two plane circular metal plates A and B, each 25 cm. in diameter, are supported 2 cm. apart. The potential difference between the plate A and the ground is indicated by a quadrant electrometer which gives a deflection of 1000 mm. for one volt. If the potential of the plate B is changed from 0 to 1 volt, the electrometer deflection changes 200 mm. If the plate B is kept at − 100 volts, A being at zero potential, and B is illuminated for 10 sec. by ultra-violet light, the electrometer gives a deflection of 100 mm. Calculate the current between the plates due to the light. Also calculate the apparent capacity of the electrometer. Ans. Current, $1{\cdot}08 \times 10^{-12}$ amp., capacity 78 cm.

5. How many quanta per second of frequency $\nu = 3 \times 10^{14}$ per second falling normally on a black plate would give a force on the plate of 1 dyne? If the plate had a mass of 1 gm. and specific heat 0·1 calorie per gramme per degree Centigrade, at what rate would its temperature rise? Ans. $1{\cdot}53 \times 10^{22}$; $7{\cdot}16 \times 10^{3}$ degrees Centigrade per second.

CHAPTER V

1. Show that the number of ways in which N like objects can be arranged in n boxes is equal to n^N. Hence show that when a gas containing N atoms expands freely from a volume V_1 to a volume V_2 its entropy is increased by $kN\log(V_2/V_1)$.

2. The energy of a collection of $\mathfrak{N}$ like systems, each of which can have energy 0 or ε only, is equal to $\frac{1}{2}\mathfrak{N}\varepsilon$. Calculate the temperature of the collection.

3. If N objects are distributed at random among M boxes, show that, when M and N are very large, the fraction of the boxes which get n objects is equal to $\dfrac{M}{N+M}\left(\dfrac{N}{N+M}\right)^n$. Show that this result agrees with the equilibrium distribution of energy among $\mathfrak{N}$ systems, each of which can have energies 0, ε, 2ε, 3ε, ...

4. Deduce Maxwell's law for the distribution of the velocities among the molecules of a gas from the quantum theory of a monatomic gas.

5. Calculate the vapour pressure of mercury at 0° C., in millimetres of mercury, by means of the equation (p. 75)

$$\log p = -\frac{L}{kN\theta} + \log\left\{\left(\frac{2\pi m}{h^2}\right)(k\theta)\right\}.$$

6. Show that, according to Debye's theory of specific heats, the shortest possible wave-length in a solid is nearly equal to the mean distance between the centres of two adjacent atoms.

7. Assuming that heat radiation consists of quanta of energy $h\nu$ and that the pressure p on the walls of a vessel containing radiation of energy density E is equal to $E/3$, show that $p = 0{\cdot}90\ldots Nk\theta$, where N is the number of quanta per cubic centimetre, k the gas constant for 1 molecule, and θ the absolute temperature.

8. If an electron of mass m and charge e is describing a circular orbit about a fixed positive charge E, and if the angular momentum of the electron can only be equal to $nh/2\pi$, where $n = 1, 2, 3, 4, \ldots$, show that the energy is equal to $A - (2\pi^2e^2E^2m)/(n^2h^2)$, where A is a constant.

9. If the acceleration of a particle relative to fixed axes is $\mathbf{A}$, show that

$$\mathbf{A} = \frac{d^2\mathbf{R}}{dt^2} + 2\left[\boldsymbol{\omega}\frac{d\mathbf{R}}{dt}\right] + \left[\boldsymbol{\omega}\,[\boldsymbol{\omega}.\mathbf{R}]\right],$$

where $\mathbf{R}$ is the distance of the particle from the origin relative to axes rotating with constant angular velocity $\boldsymbol{\omega}$.

10. For a pendulum consisting of a particle of mass m, suspended by a thread of length l, describing a circle in a horizontal plane, show that according to the quantum theory the possible values of r, the radius of the circle, are given by

$$\frac{r^8}{l^2 - r^2} = \frac{n^4h^4}{16\pi^4m^4g^2},$$

where $n = 0, 1, 2, 3, \ldots$ Show also that if l is slowly made shorter the kinetic energy divided by the frequency will remain constant.

11. For a particle of mass m moving along the x axis and acted on by a force equal to $(-\mu x)$, Schrodinger's equation $\Delta\psi + \frac{8\pi^2m}{h^2}(h\nu - V)\psi = 0$ becomes $\frac{\partial^2\psi}{\partial x^2} + \frac{8\pi^2m}{h^2}(h\nu - \frac{1}{2}\mu x^2)\psi = 0$. Transform this into $\frac{\partial^2\psi}{\partial q^2} + (C - q^2)\psi = 0$, where $C = 2\nu/\nu_0$ and $\nu_0 = \frac{1}{2\pi}\sqrt{\frac{\mu}{m}}$. Find a solution in the form $\psi = e^{-\frac{1}{2}q^2}\sum_0^\infty a_nq^n$, and show that if $C = 2n + 1$ ($n = 0, 1, 2, 3, \ldots$) the series terminates.

12. Draw a graph showing the number of electrons with momenta between p and $p + 1$ as a function of p for an electron gas with 10^{23} electrons per cm.³ at temperature 1000° K.

13. Show that the constant β in the Fermi-Dirac theory of an electron gas is equal to $1/k\theta$.

14. Show that the path of an electron in an electric field is the same as that of a ray of light in a medium of refractive index $\sqrt{(E - V)/E}$, where E is the energy of the electron and V its potential energy.

15. Show that $w = A(E - V)^{-1/4}e^{\pm\int\frac{2\pi i p dx}{h}}$, where $p = \sqrt{2m(E - V)}$, is an approximate solution of Schrödinger's equation for a particle moving along the x axis, provided that V varies slowly with x.

16. If $V = 0$ for $x < 0$ and $V = V_1$, a constant, for $x > 0$, find solutions of Schrödinger's equation $\partial^2w/\partial x^2 + 8\pi^2m(E - V)w/h^2 = 0$ which satisfy the

boundary conditions that w and $\partial w/\partial x$ must be continuous at $x = 0$. Consider the two cases $E > V_1$ and $E < V_1$.

17. Show that the normalized proper functions for the hydrogen atom with $n = 2$ are

$$w = \frac{1}{2\sqrt{2\pi a^3}} e^{-r/2a}\left(\frac{r}{2a} - 1\right), \quad w = \frac{1}{2\sqrt{2\pi a^3}} e^{-r/2a} \frac{r}{2a} P,$$

where P is $\cos\theta$, $\frac{1}{\sqrt{2}}\sin\theta e^{i\phi}$, or $\frac{1}{\sqrt{2}}\sin\theta e^{-i\phi}$ and $a = h^2/4\pi^2 me^2$.

18. Show that Schrodinger's equation for a molecule consisting of two atoms with masses m_1 and m_2 at a fixed distance d apart is

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial w}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 w}{\partial\varphi^2} + \frac{8\pi^2 IE}{h^2} w = 0,$$

where θ and φ are polar co-ordinates of the line joining the two atoms, and I is the moment of inertia of the molecules about an axis through its centre of mass and perpendicular to the line joining the two atoms.

19. Show that E_n for the molecule of (18) is given by

$$E_n = \frac{h^2}{8\pi^2 I} n(n+1), \text{ where } n = 0, 1, 2, \ldots :$$

20. Calculate the chance of a transition from w_k to w_q for the molecule of (18).

21. Show that scattering of electrons by a positively charged nucleus may be explained by supposing that the de Broglie waves of an electron are scattered by each element of volume near the nucleus and that the waves scattered have amplitude equal to $2\pi mV/h^2 r\,dx\,dy\,dz$, where V is the potential energy of an electron at the element of volume $dx\,dy\,dz$, and r is the distance from the element of volume.

22. Show that $qp - pq = ih/2\pi$, where p is the momentum belonging to the co-ordinate q. Use $pw = \frac{h}{2\pi i}\frac{\partial w}{\partial q}$, where w is a function of q, so that qpw is not equal to pqw.

23. Show that $fp_x - p_x f = \frac{ih}{2\pi}\frac{\partial f}{\partial x}$, where f is a function of x.

CHAPTER VI

1. Calculate the potential difference in volts through which an electron must drop in order to have enough energy to move the electron in a hydrogen atom from the orbit with quantum number 2 to that with number 3.

2. Calculate the average kinetic energy in ergs of electrons from a wire at 2000° C. What potential difference in volts would be required to stop an electron having this average energy?

3. The potential difference between two parallel plates is equal to V. The space between them contains a gas at a low pressure, the molecules of which are ionized by electrons which have fallen through A volts, but collisions with slower electrons are perfectly elastic. If the negative plate emits n electrons, show that the charge received by the positive plate is equal to $ne \cdot 2^p$, where p is equal to the integral part of V/A, assuming there is no appreciable recombination.

4. Calculate the ionization potential of an atom consisting of a heavy sphere of electricity of uniform density, charge e, and radius a, with an electron with charge $-e$ at the centre. What value of a would make the ionization potential V agree with that given by the equation $Ve = h\nu$, where ν is the frequency of the electron?

5. Calculate the ionization potentials of a helium atom, assuming it to consist of a heavy nucleus with charge $2e$ and two electrons moving round a circle at opposite ends of a diameter, the angular momentum being equal to h/π. When the atom has lost one electron, assume that the other one moves in a circle with angular momentum $h/2\pi$.

CHAPTER VII

1. An X-ray tube uses 30 milliamperes at 100,000 volts. Assuming that 1 per cent of the energy goes into X-rays which are emitted equally in all directions over a hemisphere, calculate the energy of the rays falling on 1 sq. cm. per second at a distance of 1 m. from the anti-cathode.

2. Calculate the wave-length of X-rays for which the energy of one quantum is 10^{-6} erg.

3. A crystal consists of equal atoms arranged on a cubical space lattice, the cubical elements having sides 5×10^{-8} cm. long. Calculate the first order glancing angles for X-rays of wave-length 10^{-8} cm. from the (100), (210), and (112) faces.

4. The refractive index μ of a substance for light of frequency ν is given by

$$\mu^2 = 1 + \frac{1}{2}\,\frac{\nu^2}{\nu_0^2 - \nu^2},$$

where $\nu_0 = 3 \times 10^{13}$. Calculate the refractive index for X-rays of wave-length 10^{-8} cm., and the critical glancing angle for total reflection of these rays at the surface of the substance.

5. If X-rays are reflected from the surface of a crystal the refractive index of which for the rays is μ and grating space d, show that

$$n\lambda = 2d\mu\sqrt{1 - (1/\mu^2)\cos^2\theta}.$$

6. If θ_1 is the value of θ found when $n = n_1$, in the previous example, and θ_2 that with n_2, show that

$$1 - \mu = \frac{n_2^2 \sin^2\theta_1 - n_1^2 \sin^2\theta_2}{n_2^2 - n_1^2}.$$

Putting $n_1\lambda_1 = 2d \sin\theta_1$, and $n_2\lambda_2 = 2d \sin\theta_2$, and $\lambda_1 - \lambda_2 = \delta\lambda$, show that when $\delta\lambda/\lambda_1$ is small,

$$1 - \mu = \frac{2n_2^2 \sin^2\theta_1\, \delta\lambda}{(n_2^2 - n_1^2)\lambda}.$$

7. Electrons are projected with velocity v from a point in a uniform magnetic field H. The electrons all start off in nearly the same direction, which is perpendicular to the field. Show that they will all pass through a line parallel to the field, and find the position of this line.

8. Show that according to the classical theory the energy scattered per second by an electron, when in a beam of plane polarized X-rays, inside a cone of solid angle $d\omega$ making an angle θ with the direction of the electric field in the rays, is equal to $(E^2e^4 \sin^2\theta d\omega)/(4\pi m^2c^3)$, where E denotes the root mean square of the electric field in the incident rays.

Hence show that for unpolarized rays the energy per second in a cone making an angle ϕ with the direction of the rays is $(E^2e^4/8\pi m^2c^3)(1 + \cos^2\phi)d\omega$.

9. Show that it is impossible for a free electron initially at rest to completely absorb a quantum of radiation.

10. Apply A. H. Compton's theory of the scattering of X-rays to the case of a quantum $h\nu$ being scattered by an electron moving initially with velocity v in a direction making an angle θ with that in which the quantum is moving.

11. Calculate the change in wave-length when light of wave length 10^{-4} cm is scattered by free electrons in a direction perpendicular to the direction of propagation of the light. Would you expect to get this change of wave-length when light is reflected from a metallic mirror?

12. Show that the change of wave-length due to scattering of X-rays by free electrons on the quantum theory is the same as that indicated by the classical theory for scattering by electrons moving in the direction of the incident rays with a certain velocity, and calculate this velocity.

CHAPTER VIII

1. The wave numbers of a spectral series are 1857, 19,932, 25,215, 27,665, 28,997. Show that the limit of this series is approximately 32,033.

2. The value of the constant N in the formula

$$\nu = Z^2N\left(\frac{1}{n^2} - \frac{1}{m^2}\right)$$

for hydrogen is 109,678·3 and for helium 109,722·9. Show that for an atom with a very large value of Z it is 109,737·7.

3. Could lines of wave numbers 3427·5, 3808·3 be lines in the spectrum of atomic hydrogen?

4. The wave numbers in a band spectrum are given by $\nu = 20{,}000 + 1000n - 90n^2$. Calculate the wave number for the head of the band.

5. The wave numbers in a band spectrum are given by $\nu = 1000(2n - 1)$ when n is positive, and $\nu = -1000(2n + 1)$ when n is negative. Calculate the moment of inertia of the molecules emitting this spectrum.

CHAPTER IX

1. A light flexible wire lies in a plane which is perpendicular to a non-uniform magnetic field. If the tension in the wire is T, and it carries a current c, show that it lies along a possible path of an electron in the field, provided

$$\frac{T}{c} = \frac{mv}{e},$$

where m is the mass, e the charge, and v the velocity of the electron.

2. A particle of mass m and charge e is projected from the origin of co-ordinates x, y, z along the x axis with velocity v. There is a magnetic field H along the z axis. If y_l is the y co-ordinate of a point on the path of the particle at $x = l$, show that provided y/x is small

$$y_l = \frac{e}{mv}\int_0^l (l - x)H\,dx.$$

3. If instead of the magnetic field, in the previous example, there is an electric field Y along the y axis, show that

$$y_l = \frac{e}{mv^2}\int_0^l \left(\int_0^x Y\,dx\right)dx.$$

4. Show that the equations of motion of a particle of mass m and charge e in a magnetic field with x, y, z components F, G, H are $m\ddot{x} = e(H\dot{y} - G\dot{z})$, $m\ddot{y} = e(F\dot{z} - H\dot{x})$, $m\ddot{z} = e(G\dot{x} - F\dot{y})$, and show by means of these equations that the velocity of the particle remains constant.

5. If F and G are zero and there is an electric field Y along the y axis, show that

$$m\ddot{x} = eH\dot{y}, \quad m\ddot{y} = -eH\dot{x} + Ye, \quad m\ddot{z} = 0.$$

Show that a solution is given by $x = A\sin(\omega t + \alpha) + Yt/H$, $y = A\cos(\omega t + \alpha)$; and find the significance of the constants A, ω, α.

6. A particle carrying a charge e and having initial kinetic energy T passes through the radial electric field between two concentric spheres of which the outer has radius R and the inner a charge E. Show that

$$h = \left\{\frac{1}{R^2} + \left(\frac{1}{k} - \frac{1}{R}\right)^2 \tan^2\left(\frac{\varphi}{2}\right)\right\}^{-\frac{1}{2}},$$

where h is the perpendicular from the centre of the spheres to the original direction of the path, $k = -Ee/2T$, and φ is the angle through which the path is deflected.

7. The velocity v of an α-ray which has travelled a distance x is given by $v^3 = A(R - x)$. Show that the time it takes to go a distance x is $t = \frac{3}{2}A^{-\frac{1}{3}}\{R^{\frac{2}{3}} - (R - x)^{\frac{2}{3}}\}$. Calculate the force on the particle as a function of x, and as a function of t.

8. If an α-ray makes a head-on collision with a hydrogen atom at rest, and the collision is perfectly elastic, calculate the ratio of the velocity given to the hydrogen atom to the initial velocity of the α-ray.

9. If α-rays are scattered by a sheet of gold containing 10^{15} atoms per square centimetre, and the initial velocity of the rays is 2×10^9 cm. per second, what fraction of the rays will be scattered through angles between 80° and 100°?

10. Calculate the potential difference in volts required to give an α-ray a velocity of 2×10^9 cm./sec.

CHAPTER X

1. What magnetic and electric fields would be necessary in J. J. Thomson's positive ray parabola apparatus (p. 230) to give equal electric and magnetic deflections of 2 cm. with positive rays consisting of oxygen atoms carrying the protonic charge e. Take $d = 5$ cm., $l = 20$ cm., and $v = 10^9$ cm./sec.

2. Mention several possible positive rays for which $e/m = 603$ approximately.

3. In Aston's mass spectrograph (fig. 4, p. 233), if the angle between AB produced and BD is denoted by θ_0 and is supposed constant, show that for BD to be independent of the velocity of the rays we must have $\theta = 2\theta_0$.

4. In Aston's mass spectrograph (fig. 4, p. 233) show that if $y = $ BD, then $dy/y = -\theta(m/e)d(e/m)$, approximately.

5. In Dempster's method of finding e/m (p. 235), show that $dr/r = -\frac{1}{2}(m/e)\,d(e/m)$.

CHAPTER XI

1. Show that the average life of the atoms of a radioactive body which decays according to the equation $N_t = N_0 e^{-\alpha t}$ is equal to $1/\alpha$.

2. Show that the average life of the atoms of a radioactive body which disappear during an interval t is equal to

$$\frac{1 \quad (1 \quad \alpha t) e^{\ldots}}{\alpha(1 - e^{\ldots})}.$$

When t is very small, show that this reduces to $\frac{1}{2}t$.

3. Calculate the weight in grammes of 1 c. m. of radon gas at 0° C. and 760 mm.

4. One milligramme of radium is in a small cavity of radius 0·5 mm. at the centre of a sphere of radius 2 cm. and thermal conductivity 0·01. If the surface of the sphere is kept at 0° C., what will be the temperature of the radium?

5. After how many years would initially pure uranium contain 1 per cent by weight of lead?

CHAPTER XII

1. If matter consists of nothing but electric point charges attracting and repelling according to the inverse square of the distance, show that, on classical theory, there is nothing to determine a definite size for any piece of matter or atom composing the piece.

2. In collisions between two particles, of equal mass, one initially at rest and one moving with velocity v, the final directions of motion relative to the centre of mass are distributed equally in all directions. Calculate the distribution relative to fixed axes.

3. The radius of a nucleus of mass number A is $2{\cdot}5 \times 10^{-13}A^{1/3}$. Assuming the particles in the nucleus all have proper functions only appreciable in equal volumes which together make up the volume of the nucleus, calculate roughly the root mean square of the momentum per particle when the nucleus is at rest.

4. Show by means of Weizsacker's formula that the known heavier nuclei cannot disintegrate with the emission of a proton.

5. Calculate the wave-length of the γ-rays for which the target area for disintegration of a deuteron is a maximum.

6. By using the atomic weights of the seven lightest elements, find if they are stable for disintegrations with the emission of protons, neutrons, or β-rays.

CHAPTER XIII

1. The gas between two parallel plates at a distance d apart is ionized so that q positive and q negative ions are produced per cubic centimetre per second. Show that the maximum or saturation current density is equal to $dqe = I$. If the current density i is less than I, show that, assuming $n_1 = n_2$ everywhere,

$$I - i = \frac{\alpha d^3 i^2}{P^2 e(k_1 + k_2)^2}$$

approximately, where P is the potential difference between the plates and α is the coefficient of recombination.

2. In Zeleny's method of finding k (p. 282), show that, if $b - a = h$ is a small quantity, then

$$k = \frac{vh}{Xd},$$

where v is the average velocity of the gas and X the electric field strength.

3. If v is the velocity of the gas between two parallel plates at a distance d apart, and ions start from one plate and reach the other at a distance h, in the direction of v, from the starting-point, show that

$$h = \frac{1}{k}\int_0^d \frac{x\,dx}{X},$$

where X is the electric field at a distance x from the plate at which the ions start.

4. In Langevin's method of finding k (p. 283), if instead of ionizing the gas over the whole distance between the plates it is only ionized in a thin layer midway between the plates, show how the charge received by the plates will vary with the time interval t.

5. Assuming that an ion moving through a gas with velocity v is retarded by a force equal to $(-\gamma v)$, show that the velocity of the ion due to an electric field X is equal to $Xe\delta/mW$, where δ is the distance the ion moves through the gas when projected with initial velocity W.

6. According to the kinetic theory of gases the coefficient of diffusion is approximately given by $K = \frac{1}{3}\lambda V$, and the velocity of an ion due to a field X is $kX = Xe\lambda/mV$. Deduce the equation $k = KNe/RT$ (p. 287) from these equations.

7. Calculate the total number of electrons in a drop of water which falls in air at the rate of 1 mm. per second.

8. Is it possible for a charged drop of mercury which falls in air at the rate of 1 mm. per second to remain at rest in a vertical field of 5000 volts per centimetre?

9. A number N of molecules in a gas is uniformly distributed over the volume of a sphere of radius a. Calculate the mean square of the distance of these molecules from the centre of the sphere after a time t.

10. Show that if N molecules in a gas describe n free paths, the mean square of the distance of these molecules from any point is increased by $2n\lambda^2$, where λ is the mean free path.

11. The vapour pressure P of a spherical uncharged drop of radius a is greater than that of a plane surface by

$$\delta P = \frac{\rho'}{\rho}\frac{2T}{a},$$

where ρ' is the density of the vapour, ρ that of the liquid, and T the surface tension of the liquid. If the drop carries a charge E, show that

$$\delta P = \frac{\rho'}{\rho}\left(\frac{2T}{a} - \frac{E^2}{8\pi a^4}\right).$$

Hence calculate the radius of a water drop carrying the protonic charge e for which $\delta P = 0$. The surface tension of water is 80 dynes cm.

12. Air saturated with water vapour at 16° C. and 760 mm. is suddenly expanded in the ratio 1 : 1·5. Calculate the amount of water which condenses, and the final temperature.

CHAPTER XIV

1. In Townsend's apparatus (p. 303), if the slit at S is replaced by a small hole at the centre of the disc B, and the lower plate GHK is divided into a circular disc of radius a surrounded by a concentric ring, show that the fraction of the charge reaching the disc of radius a is $1 - \epsilon^{-\alpha a^2}$, where $\alpha = Ze/4pkTz_a$, and z_a is the value of z at the lower plate.

2. Assuming atoms of argon to be hard spheres and electrons to be points, calculate the diameter of an argon atom if the mean free path of a high-velocity electron in argon at 1 mm. pressure is 0·147 cm. Ans. $7{\cdot}8 \times 10^{-9}$ cm.

3. Show that when an electron is moving in a strong electric field X in a gas at pressure p, and we suppose that the velocity of the electron is reduced to a

small value by each collision, then its velocity through the gas is proportional to $\sqrt{X/p}$.

4. If α denotes the number of pairs of ions produced by an electron per centimetre as it moves in a field X in a gas at pressure p, and if we suppose that the electron moves in a straight line, show that $\alpha = (p/\lambda_1)\epsilon^{-pV/X\lambda_1}$, where λ_1 is the free path of the electron at 1 mm. pressure, and V is the potential through which the electron must drop to get enough energy to ionize a molecule. Assume that every collision at which the energy of the electron is greater than Ve gives ionization and that at every collision the velocity of the electron is reduced to zero.

5. If the electric field between two parallel plates is not uniform, owing to the presence of unequal numbers of ions and electrons, show that, assuming the electrons produce ions by collision but that the ions do not,

$$\int_0^D \alpha \epsilon^{-\int_0^x \alpha\, dx} dx = \frac{1}{1+\gamma}$$

is the condition for a continuous discharge, where γ is the number of electrons liberated at the cathode by the impact of a positive ion, α is the number of pairs of ions produced by an electron per centimetre, and D is the distance between the plates. If X is constant, show that the above condition reduces to $\epsilon^{-\alpha D} = \gamma/(1+\gamma)$. If γ is a constant, and $\alpha = pf(X/p)$, show that the condition for the potential difference to be a minimum is $pD = Vf'(V/pD)/f(V/pD)$.

CHAPTER XV

1. In the theory of the variation of the potential between parallel electrodes, with an ionized gas between them, show that if, in the layer of thickness λ_1, at the positive electrode, we assume $q - \alpha n_1 n_2 = q(1 - x/\lambda_1)$, then the potential drop V_1 at this electrode is given by $V_1^2 = \frac{32}{75}\pi i\lambda_1^3/k_2$.

2. The current in amperes through a flame between two parallel electrodes at a distance d cm. apart was given by the equation

$$V = 3 \times 10^6 Cd + 6 \times 10^{11} C^2,$$

where C = current and V = potential difference. The cross-section of the flame was about 1 sq. cm. If the velocity of the negative ions due to a field of 1 volt per centimetre was 7000 cm./sec., calculate the number of negative ions per cubic centimetre in the flame. Ans. $7{\cdot}5 \times 10^7$.

3. Calculate the thickness of the layer in which the field varies, at the cathode, in the flame of the previous example, taking the mobility of the positive ions to be 2 cm./sec. per volt/centimetre, when the current is 10^{-6} amp. and $d = 10$ cm. Ans. $\lambda_2 = 0{\cdot}63$ cm.

4. If one electrode in the flame of example 2 is coated with lime or potassium carbonate, show that the current due to 630 volts will be 210 times greater in one direction than in the other.

5. Show that $\log_{10} K_2/K_1 = 5048(V_1 - V_2)/T$, where K_2 and K_1 are the equilibrium constants for the equilibrium between the vapours of two metals and electrons at the absolute temperature T, and V_1 and V_2 are the ionization potentials of the metals in volts.

6. The ionization potentials of cæsium, rubidium, potassium, sodium, and lithium are 3·87, 4·15, 4·32, 5·11, and 5·36 volts respectively. Show that the relative numbers of atoms of these metals required to give equal conductivities to a flame at 2000° C. are approximately as 1, 4·8, 13·2, 1330, 5731, when the amounts are not very small. (The relative numbers of atoms to give equal con-

ductivities to a Bunsen flame are found experimentally to be about as 1, 4·5, 7·3, 630, 4400.)

CHAPTER XVI

1. In a uniform positive column, assuming all the ionization to be due to collisions by the electrons, and all the loss of ions due to diffusion, and the rate of loss to be inversely as the pressure and proportional to the number of electrons per cubic centimetre, show that the relation between the electric field X and the pressure p must be of the form $p^2 f(X/p) =$ constant, where $f(X/p)$ denotes a function of X/p. If $X \propto \sqrt{p}$, find $f(X/p)$.

2. If in an electric discharge through a gas the electric intensity X diminishes as the current density increases, show that the cross-section of the discharge will contract until the current density has the value for which X is a minimum.

3. The "normal" cathode fall of potential is approximately equal to the minimum sparking potential. Discuss this result.

4. The variation of the potential in a discharge tube close to the anode is similar to that near the cathode but on a much smaller scale. Why is this?

5. It has been found that a uniform positive column in a discharge tube, in a magnetic field perpendicular to its length, moves sideways in a direction perpendicular to the magnetic field and to the length of the column, with a velocity approximately proportional to the magnetic field H and inversely to the pressure p. Show that we should expect the velocity to be equal to Hv_1v_2/X, where v_1 and v_2 are the velocities of the positive and negative ions along the column, and X is the electric intensity.

CHAPTER XVII

1. If 34 pairs of ions are generated per cubic centimetre per second in the air, how many positive ions will there be per cubic centimetre? Coefficient of recombination = $3400e$, with e in electrostatic units. Ans. 1600.

2. What will be the conductivity of the air in the previous example, in electrostatic units, if $k_1 = k_2 = 1·6$ cm./sec. per volt/cm.? Ans. 0·0021.

3. If 3 cm. of rain fall, consisting of drops 0·2 cm. in diameter, each carrying a charge of $4·77 \times 10^{-10}$ electrostatic units, what will be the charge received by the earth per square kilometre. Ans. 456 e.s.u.

4. A cloud consisting of a sphere of radius 0·5 km., with its centre 4 km. above the earth's surface, discharges to the earth. If there are $8·4 \times 10^4$ drops per cubic centimetre in the cloud, each carrying a charge $4·77 \times 10^{-10}$ e.s.u., what will be the change in the vertical electric field on the earth's surface, at a point 10 km. from the point vertically below the centre of the cloud? Ans. 4 volts/cm.

5. If there were a current of 500 amp. into the earth from the air in the northern hemisphere, and a current of 500 amp. from the earth into the air in the southern hemisphere, how much work in ergs would be required to take a unit magnetic pole once round the equator? Ans. 630.

CHAPTER XVIII

1. Show that $x^2 + y^2 + z^2 - c^2t^2$ is transformed into $x'^2 + y'^2 + z'^2 - c^2t'^2$ by the transformation

$$x = x' \cosh\alpha + ct' \sinh\alpha, \quad y = y',$$

$$t = t' \cosh\alpha + \frac{x'}{c} \sinh\alpha, \quad z = z'.$$

If $\alpha = \frac{1}{2}\log\left(\frac{c+v}{c-v}\right)$, show that this transformation agrees with that of the special theory of relativity.

2. If the electric field in a ray of light travelling along the x axis is proportional to $\sin\frac{2\pi}{\lambda}(x - ct)$ to an observer at rest relative to the axes x, y, z, show that to an observer moving along the x axis with velocity v the field in the ray will be proportional to

$$\sin\frac{2\pi}{\lambda'}(x' - ct'),$$

where

$$\frac{\lambda}{\lambda'} = \frac{1 - v/c}{\sqrt{1 - v^2/c^2}}.$$

3. If an observer is moving along the x axis with a velocity v and observes the volume of a body moving along the x axis with a velocity u, show that the ratio of the volume observed to that of the body when $u = v$ is

$$\sqrt{\frac{(c^2 - u^2)(c^2 - v^2)}{(c^2 - uv)^2}}.$$

4. Considering only two co-ordinates x and t, and taking the velocity of light $c = 1$, show that changing from rectangular axes x and t to new ones, which are moving relatively to x and t with velocity v along the x axis, is equivalent to changing to oblique axes with the new x axis, x', making an angle ψ with the old x axis and the new t axis, t', making an angle ψ with the old t axis so that the angle between t' and x' is $\frac{1}{2}\pi - 2\psi$. Show that $\tan\psi = v$.

5. In example 4, show that the intercepts of the curves $x^2 - t^2 = \pm 1$ on the axes x' and t' are equal to the units of length and time in the moving rectangular axes.

6. Show that, if H_x, H_y, H_z and E_x, E_y, E_z are the components of an electromagnetic field relative to axes x, y, z, the components relative to axes x', y', z' moving along x with velocity v and coinciding with x, y, z at $t = 0$ are given by

$$H_x' = H_x, \quad H_y' = \beta\left(H_y + \frac{v}{c}E_z\right), \quad H_z' = \beta\left(H_z - \frac{v}{c}E_y\right),$$

$$E_x' = E_x, \quad E_y' = \beta\left(E_y - \frac{v}{c}H_z\right), \quad E_z' = \beta\left(E_z + \frac{v}{c}H_y\right),$$

where $\beta = \frac{1}{\sqrt{1 - v^2/c^2}}$.

7. A body of mass M moving through empty space, with a velocity V, along a straight line, loses particles of matter, as it goes along, in such a way that the particles remain at rest. Show that $\frac{d}{dt}(MV) = 0$, so that it may be said that there is no force on the body, since its momentum remains constant although its velocity increases as M diminishes. If, however, the particles continue to move with velocity components, parallel to the path of the body, equal to V, show that $dV/dt = 0$, so that the momentum of the body diminishes at the rate $\left(-V\frac{dM}{dt}\right)$, and there may therefore be said to be a retarding force acting on the body equal to the rate of change of its momentum. Criticize these statements.

8. Show that the difference w between two velocities v and u, as defined in

the special theory of relativity, is given by the equation

$$\tanh^{-1}\frac{w}{c} = \tanh^{-1}\frac{v}{c} + \tanh^{-1}\frac{u}{c}.$$

9. If $M = \frac{mv}{\sqrt{(1 - v^2/c^2)}}$, $E = \frac{mc^2}{\sqrt{(1 - v^2/c^2)}}$, $F = \frac{dM}{dt}$, prove that $Fv = \frac{dE}{dt}$; m and c being constants, and v a function of t. Interpret these relations in connexion with special relativity.

CHAPTER XIX

1. What is the significance of the so-called absolute determinations of the velocity of rotation of the earth by means of Foucault's pendulum or similar devices, in the general theory of relativity?

2. If we regard the earth as at rest, which according to the principle of relativity is allowable, how can we explain the rotation round the earth of distant stars in circular orbits with velocities enormously greater than that of light?

3. If in rectangular co-ordinates x, y, z, t

$$ds^2 = -dx^2 - dy^2 - dz^2 + dt^2,$$

show that in a system of rectangular axes x_1, x_2, x_3, x_4, rotating with angular velocity ω in the x, y plane,

$$ds^2 = -dx_1^2 - dx_2^2 - dx_3^2 + \{1 - \omega^2(x_1^2 + x_2^2)\}dx_4^2 + 2\omega x_2 dx_1 dx_4 - 2\omega x_1 dx_2 dx_4.$$

4. Show directly that the equations

$$\frac{\partial^2 x_\sigma}{\partial s^2} + \{\alpha\beta, \sigma\}\frac{\partial x_\alpha}{\partial s}\frac{\partial x_\beta}{\partial s} = 0 \quad (\sigma = 1, 2, 3, 4)$$

must be the equations of motion of a particle in a gravitational field.

5. Show that Einstein's law of gravitation may be regarded as an attraction inversely as the square of the distance together with a very small attraction inversely as the fourth power of the distance.

6. Show that the deflection of a particle moving with the velocity of light past the sun on Newton's law of gravitation is one-half that indicated by Einstein's theory.

7. Show that the deflection of light by the sun, on Einstein's theory, is the same as if the space around the sun had a refractive index $1 + 2m/r$, where m is the mass of the sun and r the distance from its centre.

APPENDIX I

PERIODIC TABLE OF THE ELEMENTS

	I	II	III	IV	V	VI	VII	VIII
1	1 H 1 0078							2 He 4 002
2	3 Li 6 940	4 Be 9 02	5 B 10 82	6 C 12 00	7 N 14 008	8 O 16 0000	9 F 19 000	10 Ne 20 183
3	11 Na 22 997	12 Mg 24 32	13 Al 26 97	14 Si 28 06	15 P 31 02	16 S 32 06	17 Cl 35 457	18 A ⟵ 39 944
4	19 K 39 096 29 Cu 63 57	20 Ca 40 08 30 Zn 65·38	21 Sc 45·10 31 Ga 69 72	22 Ti 47·90 32 Ge 72 60	23 V 50 95 33 As 74 91	24 Cr 52 01 34 Se 78 96	25 Mn 54 93 35 Br 79 916	26 Fe 55 84; 27 Co 58 94 ⟵⟶ 28 Ni 58 69 36 Kr 83 7
5	37 Rb 85·44 47 Ag 107 880	38 Sr 87 63 48 Cd 112 41	39 Y 88 92 49 In 114 76	40 Zr 91 22 50 Sn 118 70	41 Cb 93 3 51 Sb 121 76	42 Mo 96 0 52 Te 127 61	43 Ma ⟵⟶ 53 I 126 92	44 Ru 101 7; 45 Rh 102 91; 46 Pd 106 7 54 Xe 131 3
6	55 Cs 132 91 79 Au 197 2	56 Ba 137 36 80 Hg 200 61	57 La 138 92 81 Tl 204 39	72 Hf 178 6 82 Pb 207·22	73 Ta 181 4 83 Bi 209 00	74 W 184 0 84 Po (210 0)	75 Re 186 31 85 —	76 Os 191 5; 77 Ir 193 1; 78 Pt 195 23 86 Rn 222
7	87 —	88 Ra 225 97	89 Ac (227)	90 Th ⟵⟶ 232 12	91 Pa (231)	92 U 238 14	93 Np 239	94 Pu 239; 95 Am 241; 96 Cm 241

THE RARE EARTHS (to be inserted between 57 La and 72 Hf)

58 Ce 140 13	59 Pr 140 92	60 Nd 144 27	61 —	62 Sm 150 43	63 Eu 152 0	64 Gd 157 3
65 Tb 159 2	66 Dy 162 46	67 Ho 163 5	68 Er 167 64	69 Tm 169 4	70 Yb 173 04	71 Lu 175 0

The numbers in front of the symbols of the elements denote the atomic numbers, the numbers underneath are the atomic weights The latter are taken, with a few modifications, from the Report of the International Commission on Atomic Weights for 1932 The double arrow ⟵⟶ indicates the places where the order of atomic weights and that of atomic numbers do not agree

APPENDIX II

TABLE OF NUMERICAL VALUES

Velocity of light,	$2{\cdot}99776 \times 10^{10}$ cm./sec.
Gravitation constant,	$6{\cdot}66 \times 10^{-8}$.
Protonic charge,	$e = 4{\cdot}8025 \times 10^{-10}$ e.s.u.
Electronic ratio of charge to mass,	$e/m_0 = 1{\cdot}7592 \times 10^{7}$ e.m.u./gm.
Faraday,	9649 e.m.u.
Planck's constant,	$h = 6{\cdot}6242 \times 10^{-27}$ erg sec.
Gas constant,	$R = 82{\cdot}05$ cm.^{3} atm. deg.$^{-1}$ mol.$^{-1}$
	$= 1{\cdot}986$ cal. deg.$^{-1}$ mol.$^{-1}$.
Molecules per mol,	$6{\cdot}0228 \times 10^{23}$.
Molecules per cm.^{3} in gas at 0° C. and 760 mm.,	$2{\cdot}69 \times 10^{19}$
Gas constant for one molecule,	$1{\cdot}379 \times 10^{-16}$ erg deg.$^{-1}$.
Mass of hydrogen atom,	$1{\cdot}6734 \times 10^{-24}$ gm.
Mass of electron,	$9{\cdot}107 \times 10^{-28}$ gm.
Atomic weight of hydrogen with $_8O^{16} = 16$,	1·00815
Energy of mass of atom of unit atomic weight with $_8O^{16} = 16$,	931 *MEV*.
Energy of electronic mass,	0·5109 *MEV*.

1 *XU* = 10^{-11} cm.

1 *AU* = 10^{-8} cm.

1 *MEV* = 10^{6} electron volts.

1 *EV* = 1 electron volt = $1{\cdot}602 \times 10^{-12}$ erg.

SUBJECT INDEX

NAME INDEX